

Cores of random graphs are born Hamiltonian

Michael Krivelevich*

Eyal Lubetzky†

Benny Sudakov‡

Abstract

Let $\{G_t\}_{t \geq 0}$ be the random graph process (G_0 is edgeless and G_t is obtained by adding a uniformly distributed new edge to G_{t-1}), and let τ_k denote the minimum time t such that the k -core of G_t (its unique maximal subgraph with minimum degree at least k) is nonempty. For any fixed $k \geq 3$ the k -core is known to emerge via a discontinuous phase transition, where at time $t = \tau_k$ its size jumps from 0 to linear in the number of vertices with high probability. It is believed that for any $k \geq 3$ the core is Hamiltonian upon creation w.h.p., and Bollobás, Cooper, Fenner and Frieze further conjectured that it in fact admits $\lfloor (k-1)/2 \rfloor$ edge-disjoint Hamilton cycles. However, even the asymptotic threshold for Hamiltonicity of the k -core in $\mathcal{G}(n, p)$ was unknown for any k . We show here that for any fixed $k \geq 15$ the k -core of G_t is w.h.p. Hamiltonian for all $t \geq \tau_k$, i.e., immediately as the k -core appears and indefinitely afterwards. Moreover, we prove that for large enough fixed k the k -core contains $\lfloor (k-3)/2 \rfloor$ edge-disjoint Hamilton cycles w.h.p. for all $t \geq \tau_k$.

1 Introduction

The fundamental problem of establishing the threshold for Hamiltonicity in the Erdős-Rényi random graph traces back to the original pioneering works of Erdős and Rényi introducing this model (see [13]). Following breakthrough papers by Pósa [29] and Korshunov [20], this question has been settled by Bollobás [4] and by Komlós and Szemerédi [19], who proved that if $p = (\log n + \log \log n + \omega_n)/n$ for any diverging increasing sequence ω_n then the probability that a random graph G drawn from the probability space $\mathcal{G}(n, p)$ of binomial random graphs is Hamiltonian tends to 1 as $n \rightarrow \infty$. This threshold precisely coincides with the one for having minimum degree 2 in $\mathcal{G}(n, p)$, and indeed a refined hitting time version of this result was proved by Bollobás [4] and by Ajtai, Komlós and Szemerédi [1]. They showed that in a typical random graph process $\{G_t\}_{t \geq 0}$ on n vertices (G_0 is edgeless and G_t is obtained by adding a uniformly distributed new edge to G_{t-1}), a Hamilton cycle appears *exactly* at the moment t when the minimum degree of G_t becomes 2 (an obvious prerequisite for Hamiltonicity).

With vertices of degree less than 2 being the primary obstacle for Hamiltonicity in the random graph, it was natural to instead consider the subgraph formed by (repeatedly) excluding such vertices. The k -core of a graph G , denoted here by $G^{(k)}$, is the (unique) maximal subgraph of G of minimum degree at least k (equivalently, it is the result of repeatedly deleting vertices of degree less than k). This concept was introduced in 1984 by Bollobás [4], who showed that for fixed $k \geq 3$ the k -core is

*School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email address: krivelev@post.tau.ac.il. Research supported in part by USA-Israel BSF Grant 2010115 and by grants 1063/08, 912/12 from Israel Science Foundation.

†Microsoft Research, One Microsoft Way, Redmond, WA 98052-6399, USA. Email address: eyal@microsoft.com.

‡Department of Mathematics, UCLA, Los Angeles, CA 90095, USA. Email: bsudakov@math.ucla.edu. Research supported in part by NSF grant DMS-1101185, by AFOSR MURI grant FA9550-10-1-0569 and a USA-Israel BSF grant.

typically nonempty and k -connected already at $p = C(k)/n$. Łuczak [23] established in 1987 that the threshold for the k -core to contain $\lfloor k/2 \rfloor$ edge-disjoint Hamilton cycles, plus a disjoint perfect matching for odd k , is when $p = (\frac{1}{k+1} \log n + k \log \log n)/n$. In particular, the 2-core of $\mathcal{G}(n, p)$ is Hamiltonian with high probability¹ when $p = (\frac{1}{3} \log n + 2 \log \log n + \omega_n)/n$ for any diverging increasing sequence ω_n , preceding the Hamiltonicity of the entire graph roughly by a factor of 3. However, for any fixed $k \geq 3$, the threshold for the k -core to contain a (single) Hamilton cycle remained open.

The best upper bound to date on this threshold is $(2 + o_k(1))(k+1)^3/n$ by Bollobás, Cooper, Fenner and Frieze [6] via their study of the related model $\mathcal{G}(n, m, k)$, a graph uniformly chosen out of all those on n labeled vertices with m edges and minimum degree at least k . It is well-known (and easily seen) that, conditioned on its number of vertices N and edges M , the k -core of $\mathcal{G}(n, p)$ is distributed as $\mathcal{G}(N, M, k)$, hence the results on the latter model for a suitable $M = M(N)$ carry to $G^{(k)}$ for $G \sim \mathcal{G}(n, p)$. It was shown in [6], following a previous paper of Bollobás, Fenner and Frieze [7], that for any fixed $k \geq 3$ the graph $G \sim \mathcal{G}(n, m, k)$ w.h.p. contains $\lfloor (k-1)/2 \rfloor$ edge-disjoint Hamilton cycles (plus a disjoint perfect matching when k is even) as long as $m/n > (k+1)^3$ (see the recent work [14] for an improvement in the special case $k = 3$). This can then be translated to $G \sim \mathcal{G}(n, p)$ to imply that, for any fixed $k \geq 3$, its k -core is Hamiltonian w.h.p. provided that $p > (2 + o_k(1))(k+1)^3/n$.

An obvious lower bound is the first appearance of the k -core, which on its own involves a remarkable first-order phase transition along the evolution of the random graph: Letting $k \geq 3$ and defining

$$\tau_k = \min \{t : G_t^{(k)} \neq \emptyset\} \quad (1.1)$$

to be the first time at which the random graph process gives rise to a nonempty k -core, Łuczak [24, 25] showed that $|G_{\tau_k}^{(k)}| \geq n/5000$ w.h.p. That is, in a typical random graph process, the k -core remains empty until the addition of a single edge creates a chain reaction culminating in a linear-sized core. This fascinating phenomenon triggered a long line of works (see, e.g., [9–12, 16, 17, 26–28, 30, 31]). Most notably, the tour-de-force by Pittel, Spencer and Wormald [28] determined the asymptotic threshold for its emergence to be $p = c_k/n$ for $c_k = \inf_{\lambda > 0} \lambda / \mathbb{P}(\text{Poisson}(\lambda) \geq k-1) = k + \sqrt{k \log k} + O(\sqrt{k})$, as well as characterized various properties of the core upon its creation, e.g., its size, edge density, etc.

Main results. The threshold for the Hamiltonicity of the k -core of $\mathcal{G}(n, p)$ is roughly between k/n and $2k^3/n$, as discussed above. Comparing these, it was long believed that the lower bound should be asymptotically tight, and indeed Bollobás, Cooper, Fenner and Frieze [6] conjectured (see also [30]) that their result could be pushed down to about k/n , and in particular, the k -core should be Hamiltonian w.h.p. for $p > c_k/n$. Furthermore, since the k -core is known to enjoy excellent expansion properties (being a random graph with a particularly nice degree distribution, namely, Poisson conditioned on being at least k) immediately upon its creation, one could expect the hitting time version of this result to also be true (see, e.g., [18, Chapter 5] for related questions). That is, w.h.p. the k -core should be Hamiltonian since the very moment of its birth in the random graph process, i.e., at time τ_k . Here we are able to show this stronger statement for any $k \geq 15$, as given by the following theorem.

Theorem 1. *Let $\{G_t\}$ be the random graph process on n vertices. Fix $k \geq 15$ and let τ_k be the hitting time for the appearance of a k -core in G_t . Then w.h.p. the k -core of G_t is Hamiltonian for all $t \geq \tau_k$.*

Note that the Hamiltonicity of the k -core is a *non-monotone* property, and as such, guaranteeing that $G_t^{(k)}$ is Hamiltonian for all $t \geq \tau_k$ is nontrivial even given that it is Hamiltonian at time $t = \tau_k$.

¹An event is said to occur with high probability (or w.h.p. for brevity) if its probability tends to 1 as $n \rightarrow \infty$.

This is due to the fact that the k -core dynamically grows with the addition of edges along the random graph process. With each new vertex that joins the k -core, we must immediately find a new cycle going through this vertex as well as all the other vertices of the core. This dynamic growth of the k -core poses additional challenges for the proofs, as we will now explain.

A common recipe for establishing Hamiltonicity in random graph models relies on *sprinkling* new edges, through which longer and longer paths are formed in the graph, eventually culminating in a Hamilton path that is followed by a Hamilton cycle. To show that the graph G at time T_1 is Hamiltonian via this framework, one can examine the graph at time $T_0 = T_1 - cn$, call it H , and show that it enjoys the following property: Out of all edges missing from H , a constant proportion are “good” in the sense that adding any single good edge will support the above extension of the maximum path length. This is typically established using strong expansion properties of H , and in particular, this property is retained when additional edges are added to the graph. As such, when adding cn edges (for an appropriate constant $c > 0$) along the interval (T_0, T_1) , one can collect (w.h.p.) up to n new good edges, leading to a Hamilton cycle. (A version of this framework also exists for a choice of $T_0 = T_1 - o(n)$, provided that the initial graph H satisfies additional more delicate properties.)

In our setting, the above framework faces a major obstacle due to the dynamical nature of the k -core, already present if one merely wishes to show the weaker asymptotic version of Theorem 1 (i.e., that for any fixed $\varepsilon > 0$ the k -core of $G \sim \mathcal{G}(n, p)$ is w.h.p. Hamiltonian at $p = (c_k + \varepsilon)/n$). Indeed, the very same sprinkling process that repeatedly extends our longest path simultaneously enlarges the k -core and might destroy the aforementioned properties of $H = G_{T_0}$ and foil the entire procedure.

Even more challenging is the obstacle in the way of a hitting time version (Hamiltonicity of $G_t^{(k)}$ at $t = \tau_k$), due to the first-order phase transition by which the k -core emerges. Already at time $\tau_k - 1$ the k -core is empty, thus inspecting $H = G_{T_0}$ for $T_0 < T_1 = \tau_k$ would not reveal which subgraph of G_{T_1} will host the future core (and the sought properties will clearly not hold for every possible subgraph).

The novelty in this work is a different setup for sprinkling, hinging on *conditioning on the future* at time T_0 . Roughly put, at time T_0 we condition on the subset of vertices V' that will host the k -core at time T_1 , as well as on some of the edges along the interval (T_0, T_1) , in a careful manner such that:

- (i) A constant fraction of the edges along the interval (T_0, T_1) remains unrevealed.
- (ii) The unrevealed edges are *uniformly distributed* over a (linear) subset of the edges that, all throughout the sprinkling process, contains most of the good edges.

Observe that one can establish various properties of the induced graph of H on V' (e.g., expansion) and these would now be retained throughout the sprinkling (as V' remains static in the new setup). Properties (i) and (ii) then enable one to appeal to the standard sprinkling framework for Hamiltonicity. In our context, combining this method with the classical *rotation-extension* technique of Pósa [29] reduces Hamiltonicity to the task of verifying certain properties of the induced graph on V' at time T_0 (namely, suitable expansion and the existence of a sub-linear number of cycles covering all vertices). The latter task is where we require that $k \geq 15$ for the arguments to hold (and be reasonably short), and it is plausible that a more detailed analysis of this technical part would relax this restriction.

Engineering the above setup to begin with, however, is highly nontrivial: Indeed, the identity of the vertex set of the future core V' can give away a great deal of information on the upcoming edges until time T_1 , easily biasing them away from being uniform on a given set of edges. Moreover, the set of good edges dynamically changes, and depends on the structure of the k -core in a complicated way...

The key step is to examine the set V' , conditioned to host the k -core at time T_1 , and partition it into $\mathcal{B}_k \subset V'$ and $\mathcal{C}_k = V' \setminus \mathcal{B}_k$, based on the graph H at time $T_0 = T_1 - o(n)$ in the following manner: The set \mathcal{B}_k will consist of every $v \in V'$ such that v has at least k neighbors in V' already at time T_0 . We then reveal all edges along the interval (T_0, T_1) *except* those with both endpoints in \mathcal{B}_k (giving away only the total number of such edges encountered along this interval) with the intuition being:

- (i) Every $v \in \mathcal{B}_k$ already guaranteed its inclusion in the k -core, thus the position of the unrevealed edges cannot interfere with our conditioning on V' , and consequently these edges are *uniform*.
- (ii) As the k -core is linear and $T_1 - T_0 = o(n)$, we have $|\mathcal{B}_k| = (1 - o(1))|V'|$ and so almost all the good edges in our sprinkling process are going to have both of their endpoints lie within \mathcal{B}_k .

Of course, turning this intuition into a rigorous argument requires a delicate analysis, working around several complications that conditioning on the future V' entails. The full details are given in Section 2.

Having settled the issue of Hamiltonicity in k -cores for $k \geq 15$, we turn our attention to *packing* Hamilton cycles, addressing the conjecture of Bollobás, Cooper, Fenner and Frieze [6] that the k -core should give rise to $\lfloor (k-1)/2 \rfloor$ edge-disjoint Hamilton cycles (plus a perfect matching when k is even).

As noted in [6], packing $\lfloor (k-1)/2 \rfloor$ edge-disjoint Hamilton cycles in the k -core of $\mathcal{G}(n, p)$ for $p = O(1/n)$ would be best possible w.h.p. The obstruction, as in the previous work of Łuczak [23], is the presence of a so-called *k-spider*, a vertex of degree $k+1$ whose neighbors all have degree k . (Whence, with k being even, packing $k/2$ Hamilton cycles would clearly use all the edges incident to these degree- k vertices, and in the process exhaust all $k+1$ edges incident to the root, impossible.) As mentioned above, packing this many Hamilton cycles (possibly plus a matching) in the k -core was shown in [6] to be possible roughly at $p = 2k^3/n$, compared with its emergence threshold of $p \sim k/n$.

Here we nearly settle this conjecture, albeit for a sufficiently large k , by extracting $\lfloor (k-3)/2 \rfloor$ edge-disjoint Hamilton cycles (one cycle short of the target number) as soon as the k -core is born. Similarly to the result in Theorem 1, this packing is then maintained indefinitely afterwards w.h.p.

Theorem 2. *Let $\{G_t\}$ be the random graph process on n vertices. Let τ_k denote the hitting time for the appearance of a k -core in G_t for some large enough fixed k . Then w.h.p. the k -core of G_t contains $\lfloor (k-3)/2 \rfloor$ edge-disjoint Hamilton cycles for all $t \geq \tau_k$.*

The proof of the above theorem follows the same framework of the proof of Theorem 1, with the main difference being the properties we require from the induced subgraph on the future k -core V' before commencing the sprinkling process. Instead of seeking a single 2-factor in this subgraph as in the proof of Theorem 1, the new goal becomes finding a $\lfloor (k-3)/2 \rfloor$ edge-disjoint 2-factors, which we can then convert to Hamilton cycles one at a time. To this end we rely on the recent results of Pralat, Verstraëte and Wormald [30] and of Chan and Molloy [10], which showed the existence of $(k-2)$ -factors and $(k-1)$ -factors in the k -core, respectively, for any sufficiently large k .

Organization and notation. In §2 we prove Theorem 1 modulo two technical lemmas (Lemmas 2.3 and 2.4, whose proofs appear in §3) on typical properties of k -cores in the critical and the supercritical regimes, resp. The proof of Theorem 2 appears in §4, and the final section §5 contains open problems.

Throughout the paper, all logarithms are in the natural basis. Our graph theoretic notation is standard: For a given graph G and subsets X, Y of its vertices, $e_G(X)$ and $e_G(X, Y)$ denote the numbers of edges of G spanned by X and between X and Y , respectively. We let $N_G(X)$ denote the set of neighbors of X in G , and occasionally refer to $N_G(X) \setminus X$ as the external neighbors of X in G .

2 Hitting time for Hamiltonicity of the core

In this section we prove Theorem 1 modulo two technical lemmas whose proofs will be postponed to Section 3. The task of establishing that the k -core of almost every random graph process is Hamiltonian as soon as it emerges and indefinitely afterwards is divided into two separate regimes: an $O(n)$ -interval of the random graph process surrounding the abrupt emergence of the k -core, which we refer to as the *critical regime*, and the $O(n \log n)$ -interval that immediately follows it, referred to as the *supercritical regime*. Our aim in the critical regime is to establish Hamiltonicity w.h.p. under the assumption that the k -core already exists, whereas in the supercritical regime we aim to maintain the Hamiltonicity property even as the k -core continues to grow. Achieving these goals are the following Theorems 2.1 and 2.2, which address the critical and supercritical regimes, respectively.

Theorem 2.1. *Fix $k \geq 15$ and let $G \sim \mathcal{G}(n, m)$ for some m satisfying $\max\{k/2, 10\}n \leq m \leq \frac{7}{6}kn$. Then with probability $1 - o(n^{-3})$ either $G^{(k)}$ is empty or it is Hamiltonian.*

Theorem 2.2. *Fix $k \geq 15$, let $H \sim \mathcal{G}(n, m)$ for $\frac{1}{6}(7k - 1)n \leq m \leq n \log n$, and let G be the graph obtained by adding $\lfloor n/8 \rfloor$ uniformly distributed new edges to H . Then the probability that $H^{(k)}$ is Hamiltonian, and at the same time $G^{(k)}$ is connected but not Hamiltonian, is at most $o(n^{-3})$.*

Using well-known properties of the k -core, Theorem 1 is obtained as an immediate corollary of Theorems 2.1 and 2.2. Indeed, taking a union bound over the interval $I_1 = [\max\{k/2, 10\}n, \frac{7}{6}kn]$ we can extend the conclusion of Theorem 2.1 to hold with probability $1 - o(n^{-2})$ simultaneously for all G_t with $t \in I_1$. Thus, provided that τ_k falls within this interval, this implies Hamiltonicity upon creation and up to $t \leq \frac{7}{6}kn$. The explicit threshold in $\mathcal{G}(n, p)$ for the sudden emergence of the k -core for $k \geq 3$ is known (see, e.g., [24, 28, 30]) to be c_k/n for $c_k = k + \sqrt{k \log k} + O(\sqrt{k})$. Specifically in our context, one can verify that $c_{15} \approx 20.98$ and $\max\{20, k\} \leq c_k \leq 2k$ holds for any $k \geq 15$. The corresponding threshold τ_k for the emergence of the k -core along the random graph process is concentrated around $c_k n/2$, and moreover it is known ([17, Theorem 1.4]) that $(\tau_k - c_k n/2)/\sqrt{n}$ converges in distribution to a Gaussian $N(0, \sigma_k^2)$ for some (explicit) $\sigma_k^2 > 0$. In particular, $\tau_k \in [\max\{k/2, 10\}n, kn]$ except with probability exponentially small in n , and altogether G_t is w.h.p. Hamiltonian for all $\tau_k \leq t \leq \frac{7}{6}kn$.

In order to maintain this property throughout the evolution of the random graph process (in which the k -core grows and consequently might not remain Hamiltonian), it suffices to do so up to $t = n \log n$, as it is well-known that $\mathcal{G}(n, m)$ for $m = \frac{1+\varepsilon}{2}n \log n$, for any constant $\varepsilon > 0$, is w.h.p. k -connected and Hamiltonian (see, e.g., [5, 18] and the references therein). A union bound over the time interval $I_2 = [\frac{7}{6}kn - n/6, n \log n]$ extends the conclusion of Theorem 2.2 to hold with probability $1 - n^{-2+o(1)}$ simultaneously for all G_t with $t \in I_2$. With this probability we have that Hamiltonicity at time $t \in I_2$ is carried to time $t + n/8$, unless the k -core at this later time happens to be not connected. It was shown by Łuczak ([25, Theorem 5'], see also [24]) that w.h.p. throughout the entire random graph process $\{G_t\}$, whenever the k -core is nonempty it is necessarily k -connected and contains at least $n/5000$ vertices. It therefore follows that Theorem 2.2 can extend the Hamiltonicity from an interval $[a, a + n/8]$ for some $a \geq \frac{7}{6}kn - n/6$ to hold w.h.p. for all $t \geq a$. As $\tau_k \leq kn$ w.h.p., with the above conclusion from Theorem 2.1 we can take $a = \lceil \frac{7}{6}kn - n/6 \rceil$ and thus G_t is w.h.p. Hamiltonian for all $t \geq \tau_k$, concluding the proof of Theorem 1.

In the remainder of this section we prove Theorems 2.1 and 2.2 modulo Lemmas 2.3 and 2.4, which are stated next and whose proofs appear in Sections 3.1 and 3.2, respectively.

Lemma 2.3. Fix $k \geq 15$ and let $G \sim \mathcal{G}(n, m)$ for $m = cn/2$ such that $\max\{k, 20\} \leq c \leq \frac{7}{3}k$. With probability $1 - o(n^{-3})$ the following holds for any $G' = (V', E') \subset G$ of minimum degree $\delta(G') \geq k$ such that $|V'| \geq \frac{4}{5}n$ and the induced subgraph of G on V' contains at most $|E'| + n/\log \log n$ edges:

- (i) the graph G' is connected,
- (ii) every $X \subset V'$ of size $|X| \leq n/c^2$ satisfies $|N_{G'}(X) \setminus X| > 2|X|$, and
- (iii) there exists a factor in G' consisting of a path P and at most $n/\sqrt{\log n}$ cycles.

Lemma 2.4. Fix $k \geq 15$, let $H \sim \mathcal{G}(n, m)$ for $\frac{1}{6}(7k-1)n \leq m \leq n \log n$ and let G be obtained by adding $n/8$ uniformly chosen new edges to H . Let $G' = (V', E') \subset G$ be the graph where $V' = V(G^{(k)})$ and $uv \in E'$ for $u, v \in V'$ if either $uv \in E(H)$ or alternatively $uv \in E(G) \setminus E(H)$ and at least one of u, v belongs to $V' \setminus V(H^{(k)})$. Then with probability $1 - o(n^{-3})$ we have $|V'| \geq |V(H^{(k)})| > 0.999n$ and every subset $X \subset V'$ of size $|X| \leq n/5$ satisfies $|N_{G'}(X) \setminus X| > 2|X|$.

Proof of Theorem 2.1. Given two graphs $H \subset G$ on the same vertex set V , define the partition of the vertex set of $G^{(k)}$, the k -core of G , into $\mathcal{B}_k = \mathcal{B}_k(G, H)$ and $\mathcal{C}_k = \mathcal{C}_k(G, H)$ as follows:

$$\mathcal{B}_k = \left\{ u \in V(G^{(k)}) : \left| N_H(u) \cap V(G^{(k)}) \right| \geq k \right\}, \quad \mathcal{C}_k = V(G^{(k)}) \setminus \mathcal{B}_k. \quad (2.1)$$

That is, \mathcal{B}_k consists of all vertices of $G^{(k)}$ which possess at least k neighbors within $V(G^{(k)})$ already in the subgraph H . It is easily seen that \mathcal{C}_k is empty iff $V(H^{(k)}) = V(G^{(k)})$, and that every vertex of \mathcal{C}_k must have an edge incident to it in $E(G) \setminus E(H)$. We will be interested in settings where $G^{(k)}$ is nonempty and $|E(G) \setminus E(H)| = o(n)$, whence \mathcal{B}_k will comprise the bulk of the k -core of G .

Let $\mathcal{S} = (e_1, \dots, e_m)$ be a uniformly chosen ordered subset of m edges out of the $\binom{n}{2}$ possible ones on the vertex set $V = [n]$, and let G and H be the graphs on the vertex set V with edge sets

$$\begin{aligned} E(G) &= \{e_1, \dots, e_m\}, \\ E(H) &= \{e_1, \dots, e_{m'}\} \quad \text{where} \quad m' = m - \frac{n}{\log \log n}. \end{aligned}$$

As $G \sim \mathcal{G}(n, m)$, if $G^{(k)}$ is empty then there is nothing left to prove. We therefore assume otherwise and aim to establish that $G^{(k)}$ is Hamiltonian with probability $1 - o(n^{-3})$. The key to the proof will be to examine the random variables $\mathcal{B}_k(G, H)$ and $\mathcal{C}_k(G, H)$ as defined above. Already we see that $|\mathcal{C}_k| \leq 2(m - m') = 2n/\log \log n$, which would make it negligible compared to the linear size of $G^{(k)}$.

Denote the indices of the subset of the final $m - m'$ edges having both endpoints in \mathcal{B}_k by

$$\mathcal{J} = \{m' < j \leq m : e_j = (u, v) \text{ for some } u, v \in \mathcal{B}_k\}, \quad (2.2)$$

and let $J = |\mathcal{J}|$. We claim that upon conditioning on $\mathcal{B}_k, \mathcal{C}_k$ (such that $\mathcal{B}_k \cup \mathcal{C}_k \neq \emptyset$ by our assumption) as well as \mathcal{J} and all the edges $\mathcal{S}^* = \{e_j : j \notin \mathcal{J}\}$, the remaining J edges of \mathcal{S} are uniformly distributed over all edges missing from \mathcal{B}_k (that is, edges with both endpoints in \mathcal{B}_k that did not appear among $e_1, \dots, e_{m'}$). To establish this it suffices to show that adding any possible set of J new edges with endpoints inside \mathcal{B}_k to \mathcal{S}^* would result in a graph such that $\mathcal{B}_k \cup \mathcal{C}_k$ is consistent with our conditioning (since each such extension would have the same probability, and the partition $V' = \mathcal{B}_k \cup \mathcal{C}_k$ only depends on the graph H and the vertex set V').

Indeed, letting $F = (V, \mathcal{S}^*)$ we argue that $V(F^{(k)}) = V(G^{(k)})$. The fact that $V(F^{(k)}) \subset V(G^{(k)})$ follows trivially from having $\mathcal{S}^* \subset \mathcal{S}$. At the same time, $V(G^{(k)}) \subset V(F^{(k)})$ since every $u \in \mathcal{C}_k(G, H)$

has $N_F(u) = N_G(u)$ (which contains, in particular, at least k vertices from $V(G^{(k)})$) whereas every $u \in \mathcal{B}_k(G, H)$ has $N_H(u) \subset N_F(u)$ and $|N_H(u) \cap V(G^{(k)})| \geq k$ by the definition of $\mathcal{B}_k(G, H)$. It is now clear for any subset S of edges whose endpoints all lie within $\mathcal{B}_k(G, H)$ (or even within $V(G^{(k)})$ for that matter), the vertex set of the k -core of $F' = (V, \mathcal{S}^* \cup S)$ remains equal to $V(G^{(k)})$. That is, every extension of \mathcal{S}^* via J new internal edges of \mathcal{B}_k is consistent with our conditioning, as claimed.

A delicate point worth noting is that the above statement would be false if we were to directly take $m = \tau_k$, the hitting time for the emergence of the k -core. In that case, it could occur that adding certain edges between the vertices of \mathcal{B}_k to the graph $F = (V, \mathcal{S}^*)$ would induce a different k -core to appear at some earlier time $j < m$, thus violating our conditioning. This issue disappears once we fix m deterministically, thereby guaranteeing that no edges within \mathcal{B}_k would be forbidden.

The next crucial step is controlling the random variable J , which counts the number of edges within \mathcal{B}_k between times $m' + 1, \dots, m$. Our goal is to show that $J/(m - m')$ is uniformly bounded from below, that is, that a constant fraction of the edges $e_{m'+1}, \dots, e_m$ are internal edges of \mathcal{B}_k . Intuitively, one could hope that the distribution of the edges in $E(G) \setminus E(H)$ would be close to uniform. In that case, since the graphs G, H under consideration are such that \mathcal{B}_k is of linear size, standard concentration arguments would yield the sought estimate except with a probability that is exponentially small in n . While such an argument would be valid for a *fixed* set of vertices in H , unfortunately the set \mathcal{B}_k is random and *does* depend on $e_{m'+1}, \dots, e_m$.

For instance, consider the situation where conditioned on H and then on the partition $\mathcal{B}_k \cup \mathcal{C}_k$ we have that $\mathcal{C}_k = 2(m - m')$. This occurs when the edges $e_{m'+1}, \dots, e_m$ form a matching on the vertices of \mathcal{C}_k (so as to accommodate the degree constraints of all $2(m - m')$ of its vertices). Rare as this event may be, it demonstrates the possibility that $J = 0$ (even deterministically) in this conditional space. Another example of the delicate dependence between J and \mathcal{B}_k is the following: Suppose that instead of \mathcal{B}_k we would work with its variant \mathcal{B}'_k , consisting of all vertices of $V(G^{(k)})$ whose set of neighbors within $V(G^{(k)})$ is the same in G as it is in H . Then $\mathcal{B}'_k \subset \mathcal{B}_k$ and these two sets differ by at most $m - m' = o(n)$ vertices, and yet the variable J' corresponding to \mathcal{B}'_k satisfies $J' = 0$ by definition.

As the next claim shows, it is possible to control the random variable J despite the dependence between the \mathcal{B}_k and $E(G) \setminus E(H)$.

Claim 2.5. *Let $J = |\mathcal{J}|$ for \mathcal{J} as defined in (2.2). Then*

$$\mathbb{P}(J \geq n/(100 \log \log n)) = 1 - O(n^{-9}).$$

Proof. The idea underlying this approach would be to approximate \mathcal{B}_k via another random set whose identity is completely determined by G (as opposed to the combination of G and H).

Let $A_1 = A_1(G)$ denote the event that there is a subset S of $s = \lfloor \frac{2}{5}n \rfloor$ vertices of G which spans at most $r = m/50$ edges. Taking $p = c/n$ for $c = 2m/n$ we appeal to the standard correspondence between $\mathcal{G}(n, m)$ and $\mathcal{G}(n, p)$, and since $r \sim \frac{1}{16}s^2p$ we get that

$$\begin{aligned} \mathbb{P}(A_1) &\leq (r+1) \binom{n}{s} \binom{\binom{s}{2}}{r} p^r (1-p)^{\binom{s}{2}-r} \leq \left((1+o(1)) \frac{en}{s} \left(\frac{es^2p}{2r} \right)^{r/s} e^{-sp/2} \right)^s \\ &\leq \left(\left(\frac{5}{2}e + o(1) \right) (8e)^{c/40} e^{-c/5} \right)^s = \left(\left(\frac{5}{2}e + o(1) \right) (8e^{-7})^{c/40} \right)^s < (3/4)^s, \end{aligned}$$

where the last inequality uses our hypothesis that $c \geq 20$. It follows that the probability of encountering A_1 is at most $O(\exp(-a_0 n))$ for some absolute $a_0 > 0$.

Let $X_l = X_l(G)$ denote the number of vertices of degree precisely l in $G^{(k)}$ for a fixed integer $l \geq k$. It is known ([9, Corollaries 2 and 3]) that the fraction of vertices of degree l within the k -core converges to $\mathbb{P}(Z = l)$ where Z is a Poisson random variable with mean μ for some explicitly given $\mu = \mu(k) > k$. More precisely, for any fixed $\varepsilon > 0$,

$$\mathbb{P}\left(\left|X_l - \frac{e^{-\mu}\mu^l}{l!}n\right| > \varepsilon n\right) = O(e^{-n^{a_1}}), \quad (2.3)$$

where $a_1 = a_1(k) > 0$ is fixed. One should note that [9] offers (sharper) estimates only for the supercritical range $c > c_k$. However, one may extend these to $c \sim c_k$, given that the k -core is nonempty at that point, as follows. Fixing some arbitrarily small $\delta > 0$, at $c' = c_k + \delta$ these bounds yield concentration for X_l within a window of width εn (for any fixed $\varepsilon > 0$) except with an exponentially small probability. Next, appealing to the known estimates on $|V(G^{(k)})|$ and $|E(G^{(k)})|$ throughout the critical window for the emergence of the k -core (see [17, Theorem 1.3]), it is known that the difference in these two random variables between times $cn/2$ and $c'n/2$ is at most some εn except with probability $O(e^{-n^{a_1}})$ for some $a_1 > 0$ fixed, where ε can be made arbitrarily small via selecting a suitably small δ . Consequently, the value of X_l cannot change by more than $2\varepsilon n$ along this interval (accounting for lost edges as well as lost vertices via the change in $V(G^{(k)})$ and $E(G^{(k)})$, respectively), thus establishing (2.3) with a window of $3\varepsilon n$. To complement this bound, observe that for fixed $l \geq 15$,

$$\frac{e^{-\mu}\mu^l}{l!} \leq \frac{e^{-\mu}\mu^l}{\sqrt{2\pi l}(l/e)^l} = \frac{e^{-(\mu-l)}(1 + \frac{\mu-l}{l})^l}{\sqrt{2\pi l}} \leq \frac{1}{\sqrt{2\pi l}} < \frac{1}{8},$$

where the first inequality above followed from Stirling's formula and the last one holds for $l \geq 15$. Altogether, we can infer that the number of vertices whose degree belongs to $\{k, k+1, k+2\}$ satisfies

$$\mathbb{P}(X_k + X_{k+1} + X_{k+2} \leq \frac{2}{5}n) = 1 - O(e^{-a_1 n}).$$

At the same time, known properties of the k -core upon its emergence (recall that $G^{(k)}$ is nonempty by assumption), namely the explicit formula for its typical initial size as well its concentration around its mean (see, e.g., [28, Theorem 3]), imply that for some absolute $a_2 > 0$ and any fixed $k \geq 15$,

$$\mathbb{P}\left(|V(G^{(k)})| \geq \frac{4}{5}n\right) = 1 - O(e^{-n^{a_2}}).$$

Let $A_2 = A_2(G)$ denote the event that $X_k + X_{k+1} + X_{k+2} > \frac{2}{5}n$ or $|V(G^{(k)})| < \frac{4}{5}n$. The last two estimates then imply that $\mathbb{P}(A_2) \leq O(e^{-n^{a_2}})$, and on the event A_2^c we see that at least $\frac{2}{5}n$ vertices have degree at least $k+3$ in $G^{(k)}$.

A final ingredient we need is some control over vertices of large degree in G . Set

$$\kappa_1 = \log \log n, \quad \kappa_2 = 10 \frac{\log n}{\log \log n}, \quad \rho = \frac{n}{\log^2 n}, \quad (2.4)$$

and let $A_3 = A_3(G)$ denote the event that either $\Delta(G) \geq \kappa_2$, where $\Delta(G)$ denotes the maximum degree of G , or there are ρ vertices of G whose degree exceeds κ_1 . It is well-known that whenever m/n is uniformly bounded from above, the maximum degree in $\mathcal{G}(n, m)$ is at most $(1 + o(1)) \frac{\log n}{\log \log n}$ w.h.p. (see, e.g., [5, §3]), and moreover, the probability this maximum degree would exceed κ_2 as given in (2.4) is at most $O(n^{-9})$. Turning to the probability that a given $v \in V$ has degree at least κ_1 in G , again working with the corresponding $\mathcal{G}(n, p = c/n)$ model with $c = 2m/n$ we find it to be

$$\mathbb{P}(\text{Bin}(n-1, p) \geq \kappa_1) = O(c^{\kappa_1}/\kappa_1!) = e^{-(1-o(1))\kappa_1 \log \kappa_1} < e^{-5\kappa_1} = (\log n)^{-5},$$

where the strict inequality holds for large n . We will now argue that the probability of encountering ρ such vertices (recall that $\rho = n^{1-o(1)}$) would be $\exp(-n^{1-o(1)})$. Indeed, the probability of encountering a set S of ρ vertices such that the induced subgraph on S contains more than 10ρ edges is at most

$$\binom{n}{\rho} \binom{\binom{\rho}{2}}{10\rho} (c/n)^{10\rho} \leq \left(\left(\frac{en}{\rho} \right)^{\frac{1}{10}} \frac{e\rho c}{20n} \right)^{10\rho} = \left((e \log^2 n)^{\frac{1}{10}} \frac{ec}{20 \log^2 n} \right)^{10\rho} = e^{-n^{1-o(1)}}.$$

At the same time, in a set S of ρ vertices whose degrees are all at least κ_1 and where additionally the induced subgraph on S has at most 10ρ edges we must have at least $\rho(\kappa_1 - 20)$ edges in the cut between S and $V(G) \setminus S$. Similarly to the previous calculation, the probability of this event is at most

$$\binom{n}{\rho} \binom{\rho(n-\rho)}{\rho(\kappa_1-20)} (c/n)^{\rho(\kappa_1-20)} = \left(e \log^2 n \left(\frac{e(n-\rho)c}{(\kappa_1-20)n} \right)^{\kappa_1-20} \right)^{\rho} = e^{-(1-o(1))\rho\kappa_1 \log \kappa_1} = e^{-n^{1-o(1)}}.$$

Altogether we deduce that A_3 occurs with probability at most $O(n^{-9})$ with room to spare.

Let $U \subset V(G^{(k)})$ be the set of all vertices whose degree in $G^{(k)}$ is at least $k+3$. It is important to note that U as well as the properties addressed by the events A_1, A_2, A_3 are entirely determined by the edge set $E(G) = \{e_1, \dots, e_m\}$ regardless of the order in which they appeared (as opposed to $\mathcal{B}_k, \mathcal{C}_k$ which were a function of G and H). Condition on this edge set $E(G)$, unordered, and further condition on the event $A^c = A_1^c \cap A_2^c \cap A_3^c$, recalling that $\mathbb{P}(A^c) = 1 - O(n^{-9})$ by the above analysis.

Under this conditioning, the edge set $E(H)$ is obtained as a uniform subset of m' of these edges, hence the variable

$$J_U = \{uv \in E(G) \setminus E(H) : u, v \in U\}$$

is hypergeometric. Specifically, J_U is the result of $m - m' = n/\log \log n$ samples without repetition with the target population having proportion at least $1/50$, since A_1^c guarantees at least $m/50$ internal edges within any set of $\frac{2}{5}n$ vertices whereas A_2^c implies that $|U| \geq \frac{2}{5}n$. Hoeffding's inequality for hypergeometric variables [15] now tells us that the probability that for some absolute constant $a_3 > 0$,

$$\mathbb{P}(J_U < (m - m')/75 \mid E(G), A^c) \leq e^{-a_3(m-m')} = e^{-n^{1-o(1)}},$$

where the last inequality holds for large enough n .

Now let $W_1 \subset U$ denote the subset of vertices of U which have degree at most κ_1 in G and are incident to at least 4 edges of $E(G) \setminus E(H)$. The number of edges in the sample $E(G) \setminus E(H)$ that are incident to a fixed set of s such vertices is stochastically dominated by a binomial variable $\text{Bin}(m - m', s\kappa_1/m')$, as there are overall at most $s\kappa_1$ edges incident to this set in G . Hence,

$$\begin{aligned} \mathbb{P}(|W_1| \geq \rho \mid E(G), A^c) &\leq \mathbb{P}(\text{Bin}(m - m', \rho\kappa_1/m') \geq 2\rho) \leq \binom{m - m'}{2\rho} \left(\frac{\rho\kappa_1}{m'} \right)^{2\rho} \\ &\leq \left(\frac{en}{2\rho \log \log n} \frac{\rho \log \log n}{(10 - o(1))n} \right)^{2\rho} = \left(\frac{e}{20} + o(1) \right)^{2\rho} = e^{-n^{1-o(1)}}, \end{aligned}$$

where the first inequality in the second line used the hypothesis $m' \sim m \sim cn/2 \geq 10n$. Conditioned on this event, our assumption on the maximal degree of G implies that the number of edges incident to W_1 in G is at most $\kappa_2\rho = o(n/\log n)$. Similarly, letting $W_2 \subset U$ be the set of vertices whose degree in G exceeds κ_1 , we know by assumption that $|W_2| \leq \rho$ and that the number of edges incident to this set in G is then at most $\kappa_2\rho = o(n/\log n)$.

Letting $U^* = U \setminus (W_1 \cup W_2)$, by definition we have that every vertex of U^* is incident to at most 3 edges of $E(G) \setminus E(H)$. At the same time, the degree in $G^{(k)}$ of every $v \in U$ is at least $k + 3$, and consequently each $v \in U^*$ has at least k neighbors among $V(G^{(k)})$ in H , hence $U^* \subset \mathcal{B}_k$. The above estimates show that, except with probability $\exp(-n^{1-o(1)})$, there are at least $n/(75 \log \log n)$ edges with both endpoints in U (counted by the variable J_U above) whereas the total number of edges incident to $W_1 \cup W_2 = U \setminus U^*$ is $o(n/\log n) = o(n/\log \log n)$. Therefore,

$$\mathbb{P}(J < n/(100 \log \log n) \mid E(G), A^c) < e^{-n^{1-o(1)}},$$

and the desired (unconditional) estimate on J now follows from the fact that $\mathbb{P}(A^c) = 1 - O(n^{-9})$. ■

We can now turn to Lemma 2.3, taking the target graph G' to be $F^{(k)}$ for $F = (V, \mathcal{S}^*)$. As mentioned above, it is known that whenever the k -core for $k \geq 15$ is nonempty, its size is at least $\frac{4}{5}n$ except with probability exponentially small in n . Recalling that $V(F^{(k)}) = V(G^{(k)})$ and that $E(G) \setminus E(F)$ contains $J \leq n/\log \log n$ internal edges of \mathcal{B}_k , the requirements of the lemma are satisfied. This lemma will be used in tandem with the following well-known corollary of the classical rotation-extension technique of Pósa [29] (see, e.g., [21, Corollary 2.10], which is formulated slightly differently though the exact same proof implies both statements), where here and in what follows the length of paths/cycles will refer to the number of edges in them:

Lemma 2.6. *Let r be a positive integer, and let $G = (V, E)$ be a graph where for every subset $R \subset V$ with $|R| < r$ we have $|N_G(R) \setminus R| \geq 2|R|$. Then for any path P in G , denoting its length by h ,*

- (i) *there is path P' of length $h + 1$ in G containing all vertices of P plus a new endpoint $u \notin P$, or*
- (ii) *there is a cycle C' in G of length $h + 1$ on the same vertex set as P , or*
- (iii) *there are at least $r^2/2$ non-edges in G such that if any of them is turned into an edge, then the new graph contains a cycle C' as above, i.e., an $(h + 1)$ -cycle on the same vertex set as P .*

Lemma 2.3 enables us to apply the rotation-extension technique to G' with $r = n/c^2$. Let P be the path in G' such that $V(G') \setminus P$ can be covered by $\ell \leq n/\sqrt{\log n}$ cycles as per the conclusions of Lemma 2.3. Let h denote the length of P (i.e., P has h edges) and let $\{C_1, \dots, C_\ell\}$ denote the cycles covering $V(G') \setminus P$. We now argue that either G' is Hamiltonian or there exist at least $r^2/2$ edges — to be referred to as *boosters* — the addition of each of which would create either a Hamilton cycle or a path P' containing all vertices of P in addition to one of the C_i 's. Thereafter, we will sprinkle the edges e_{i_1}, \dots, e_{i_J} to repeatedly absorb all ℓ cycles into P and then form a Hamilton cycle.

To justify the above claim, examine the three possible conclusions of Lemma 2.6 above:

- (i) If the path P can be extended to a path P' ending at some vertex $u \in V(G') \setminus P$, delete an edge to unravel the cycle C_i containing u into a path, and concatenate that path to P' .
- (ii) If there is a cycle C' on the same vertex set as P , either C' is a Hamilton cycle in G' as required, or the connectivity of G' would imply that C' is connected by a path to some $u \in V(G') \setminus P$. In the latter case, unravel both C' and the cycle C_i of u in the obvious way into a single path.
- (iii) Otherwise, there are at least $r^2/2$ edges, the addition of each of which would lead to Case (ii).

Overall, letting $G'_0 = G'$ and G'_j be the result of adding the edge e_{i_j} to G'_{j-1} for $j = 1, \dots, J$, we observe that the above mentioned properties of G' are satisfied for each j (being closed under the addition of edges) and so encountering $\ell + 1 \leq n/\sqrt{\log n} + 1$ rounds for which e_{i_j} is a booster in G'_{j-1} would guarantee Hamiltonicity.

The proof can now be concluded from arguing that, for any $j = 1, \dots, J$, the probability that e_{i_j} is a booster in G'_{j-1} is uniformly bounded from below. Indeed, e_{i_j} is uniformly distributed over all missing edges in the induced subgraph of G'_{j-1} on \mathcal{B}_k . Of course, $V(G'_{j-1}) = V(G^{(k)}) = \mathcal{B}_k \cup \mathcal{C}_k$ so there are $o(n^2)$ edges incident to \mathcal{C}_k in G'_{j-1} . It follows that almost all of the boosters of G'_{j-1} are edges whose both endpoints lie in \mathcal{B}_k , and the probability that e_{i_j} belongs to this set is at least $(1 - o(1))(r^2/2)/\binom{|\mathcal{B}_k|}{2} \geq (r/n)^2 \geq 1/c^4$. Hence, conditioned on the event $J \geq n/(100 \log \log n)$, the number of boosters we collect (formally defining every new edge as a booster upon achieving Hamiltonicity) stochastically dominates the random variable $Y \sim \text{Bin}(n/(100 \log \log n), 1/c^4)$, which in turn satisfies $\mathbb{P}(Y \geq c'n/\log \log n) \geq 1 - \exp(-n^{1-o(1)})$ for some $c' > 0$ (that depends on c). In particular, $Y > n/\sqrt{\log n} + 1$ with probability $1 - \exp(-n^{1-o(1)})$ and the proof is complete. ■

Proof of Theorem 2.2. The proof will follow the same line of argument used to prove Theorem 2.1, albeit in a somewhat simpler setting as we can utilize the Hamilton cycle in $H^{(k)}$ as the foundation of our 2-factor, and furthermore, we can afford to select the subset \mathcal{B}_k of vertices — the bulk of the k -core to be used for sprinkling — based on the information of H (in lieu of conditioning on future information from the random graph process). Namely, we will choose \mathcal{B}_k to be the vertex set of $H^{(k)}$. Recall that a key element in the proof of Theorem 2.1 revolved around controlling the random variable J despite its delicate dependence on \mathcal{B}_k . As mentioned there, this issue would be circumvented should \mathcal{B}_k be determined already by graph H , precisely our present situation.

Let $H \sim \mathcal{G}(n, m)$ for $m \geq \frac{1}{6}(7k-1)n$. Ignoring floors and ceilings for brevity here and throughout this proof, let $e_1, \dots, e_{n/8}$ be a uniformly chosen subset of the edges missing from $E(H)$, and let G be the result of adding these edges to H .

Suppose that $H^{(k)}$ is Hamiltonian (otherwise there is nothing to prove), let $\mathcal{B}_k = V(H^{(k)})$, and observe that thanks to Lemma 2.4 we already know that $|\mathcal{B}_k| > 0.999n$ with probability $1 - o(n^{-3})$. To exploit the other conclusions of that lemma, we must first recover the graph G' as defined there.

As before, we will condition on the graph G' as well as on the random variable J counting the number of edges with both endpoints in \mathcal{B}_k , without revealing the identity of those edges. As we mentioned, the fact that \mathcal{B}_k is determined by H will readily yield the sought lower bound on J . Formally, for each $t = 1, \dots, n/8$, expose whether the new edge $e_t = uv$ (uniformly distributed out of all missing ones) has either $u \notin \mathcal{B}_k$ or $v \notin \mathcal{B}_k$, and if so, expose its endpoints u, v themselves. This process reveals the graph G' and random variable J , yet we see that each edge is uniformly distributed out of all missing ones (as we no longer have a conditioning involving the graph process at future times as in the proof of Theorem 2.1). At all times we have at most $m \leq n \log n + n/8$ edges in our graph (as per the scope of Theorem 2.2), therefore the probability that e_t has both its endpoints in \mathcal{B}_k is at least $((0.999n)^2 - 2n \log n) / \binom{n}{2} > 0.998$, with the last inequality being valid for large enough n . That is to say, the random variable J counting the number of such edges that we encounter stochastically dominates a binomial variable $Y \sim \text{Bin}(n/8, 0.998)$, whence standard concentration arguments imply that $J \geq n/10$ except with probability exponentially small in n . Finally, by the definition of the process above, each edge e_t that was accounted for in J is a uniform edge among all those missing from the induced subgraph on \mathcal{B}_k at the end of the previous iteration.

Applying Lemma 2.4 to $G' = (V', E')$, we find that with probability $1 - o(n^{-3})$, every subset X in V' of size $|X| \leq n/5$ has at least $2|X|$ external neighbors in G' . We claim that it is moreover connected with the same probability: Indeed, the only edges in $E(G^{(k)}) \setminus E(G')$ are edges in $E(G) \setminus E(H)$ with

both endpoints in \mathcal{B}_k , yet by assumption \mathcal{B}_k is connected already in the subgraph $H \subset G'$ since $H^{(k)}$ is guaranteed to contain a Hamilton cycle. Therefore, the fact that $G^{(k)}$ is connected carries to G' .

We are now in a position to conclude the proof in the same manner used to prove Theorem 2.1. Considering all points in $V' \setminus \mathcal{B}_k$ as trivial cycles (of which there are strictly less than $0.001n$ since $|\mathcal{B}_k| > 0.999n$) and adding those to the simple cycle that goes through all vertices of \mathcal{B}_k in $H^{(k)}$ we arrive at a 2-factor in G' with at most $0.001n + 1$ cycles. By the expansion properties of G' that were detailed above, we can derive from Lemma 2.6 that there are at least $n^2/50$ boosters in G' . However, all but $0.001n^2$ such boosters have both of their endpoints in \mathcal{B}_k , therefore the probability that a new edge, uniformly chosen out of all edges missing from the induced subgraph on \mathcal{B}_k , is a booster, is at least $1/50$ (with room to spare). Seeing as G' is connected, sprinkling any such edge to G' would increase the length of the longest path in G' , and encountering $0.001n + 1$ boosters would culminate in a Hamilton cycle through the vertices of V' . Iterating this procedure for $J \geq n/10$ rounds, the number of boosters encountered before Hamiltonicity is achieved stochastically dominates a binomial random variable $Z \sim \text{Bin}(n/10, 1/50)$. Since Z concentrates around its mean of $0.002n$ we deduce that a Hamilton cycle will be formed except with probability exponentially small in n , as desired. ■

3 Expansion and factors in cores

This section contains the proofs of Lemmas 2.3 and 2.4, which address properties of the random graph near and beyond the k -core threshold. These two regimes are studied in Sections 3.1 and 3.2, respectively, yet we begin with a straightforward fact on the expansion of small sets, to be used in both regimes.

Claim 3.1. *Fix $k \geq 15$, let $G \sim \mathcal{G}(n, p)$ for $p = \lambda/n$ with $1 < \lambda < \log^2 n$. With probability $1 - o(n^{-3})$ the following holds for any subgraph $G' = (V', E') \subset G$ (not necessarily induced) with $\delta(G') \geq k$. We have $|N_{G'}(X) \setminus X| > 3|X|$ for every $X \subset V'$ of size $|X| \leq \frac{3}{7}\lambda^{-15/7}n$, whereas $|N_{G'}(X) \setminus X| > 2|X|$ for every $X \subset V'$ of size $|X| \leq \frac{4}{5}\lambda^{-5/3}n$.*

Proof. Consider some $X \subset V'$ of size at most θn that does not expand to at least $b|X|$ external neighbors in G' for some fixed $b \geq 2$, and let $Y \subset V'$ denote the external neighbors of X in G' . The fact that $\delta(G') \geq k$ implies that out of the $\binom{|X|}{2} + |X||Y|$ potential edges incident to X , at least $k|X|/2$ edges exist in G' , hence also in G . Whenever $b \leq k/2 - 1$, the probability of this event is thus at most

$$\begin{aligned} \mathbb{P} &\leq \sum_{x \leq \theta n} \binom{n}{x} \binom{n}{bx} \left(\frac{x^2}{2} + bx^2 \right) p^{kx/2} \leq \sum_{x \leq \theta n} \left[\frac{en}{x} \left(\frac{en}{bx} \right)^b \left(\frac{(2b+1)ex}{k} \right)^{\frac{k}{2}} \left(\frac{\lambda}{n} \right)^{\frac{k}{2}} \right]^x \\ &\leq \sum_{x=1}^{\lfloor 25 \log n \rfloor} \left[n^{-k/2+b+1+o(1)} \right]^x + \sum_{x=\lceil 25 \log n \rceil}^{\lfloor \theta n \rfloor} \left[\frac{(e/\theta)^{b+1}}{b^b} ((2b+1)e\lambda\theta/k)^{\frac{k}{2}} \right]^x. \end{aligned} \quad (3.1)$$

Taking $b = 3$ and $\theta = \frac{3}{7}\lambda^{-15/7}$, the first summation is $o(n^{-3})$, whereas for $k = 15$ the occurrences of λ in the second summation cancel (and for $k > 15$ the exponent of λ will be negative), thus reducing that summation into at most

$$\sum_{x=\lceil 25 \log n \rceil}^{\lfloor \frac{3}{7}\lambda^{-15/7}n \rfloor} \left[\frac{(7e/3)^4}{27} \left(\frac{3e}{k} \right)^{\frac{k}{2}} \right]^x \leq \sum_{x=\lceil 25 \log n \rceil}^{\lfloor \frac{3}{7}\lambda^{-15/7}n \rfloor} e^{-x/3} = o(n^{-8}),$$

where we used the fact that $\frac{(7e/3)^4}{27}(3e/k)^{k/2} \leq \frac{2}{3}$ for any $k \geq 15$.

Similarly, taking $b = 2$ and $\theta = \frac{4}{5}\lambda^{-5/3}$, the first summation in (3.1) remains $o(n^{-3})$ and again we can omit all occurrences of λ in the second summation for $k \geq 15$, reducing it into at most

$$\sum_{x=\lceil 25 \log n \rceil}^{\lfloor \frac{4}{5}\lambda^{-5/3}n \rfloor} \left[\frac{(5e)^3}{256} \left(\frac{4e}{k} \right)^{\frac{k}{2}} \right]^x \leq \sum_{x=\lceil 25 \log n \rceil}^{\lfloor \frac{4}{5}\lambda^{-5/3}n \rfloor} e^{-x/8} = o(n^{-3}),$$

as required. ■

3.1 Properties of sparse random graphs near the core threshold

Throughout this subsection we will restrict our attention to $\mathcal{G}(n, p)$ in the regime $p = O(1/n)$. The next claim establishes that in any induced subgraph of minimum degree $k - 1$, sets that are large enough — namely, ones whose size is at least \sqrt{n} — have many external neighbors. This can be established up to a linear scale, e.g., for sets up to size $n/15$, solely based on an assumption that the minimum degree of the graph under consideration (a subgraph of our random graph) is at least k . However, as we will later see, it will be imperative to obtain this estimate for larger sets, to which end we will rely on additional properties that are available to us in the framework of Lemma 2.3.

Claim 3.2. *Fix $k \geq 15$ and let $G \sim \mathcal{G}(n, c/n)$ for some fixed $k \leq c \leq \frac{7}{3}k$. The following holds with probability at least $1 - O(0.8\sqrt{n})$. If $G' = (V', E')$ is an induced subgraph of G on $|V'| \geq \frac{4}{5}n$ vertices and $H = (V', E' \setminus (E_1 \cup E_2))$ where E_1 is a matching and $|E_2| \leq n/\log \log n$ such that the minimum degree of H is $\delta(H) \geq k - 1$, then every $X \subset V'$ of size $\sqrt{n} \leq |X| \leq \frac{2}{7}n$ satisfies $|N_H(X) \setminus X| > n^{1/3}$.*

Proof. First consider the range $\sqrt{n} \leq |X| \leq \theta_1 n$ for $\theta_1 = \frac{2}{23}$. Exactly as in the proof of Claim 3.1, the probability that there exists a subgraph H of G with minimum degree $k - 1$ (not necessarily formed by deleting the edge sets E_1 and E_2 as above) within which we can find a set $X \subset V$ of size x such that $\sqrt{n} \leq x \leq \theta_1 n$ and all of its external neighbors belong to some $Y \subset V$ of size $y = n^{1/3}$ is at most

$$\sum_{\sqrt{n} \leq x \leq \theta_1 n} \left[\frac{en}{x} \left(\frac{en}{y} \right)^{y/x} \left(\frac{e(x+2y)}{k-1} \right)^{\frac{k-1}{2}} \left(\frac{c}{n} \right)^{\frac{k-1}{2}} \right]^x, \quad (3.2)$$

where in comparison with (3.1) here we let b assume the role of y/x . Examining the base of the exponent in each summand we see that $(en/y)^{y/x} \leq \exp(n^{-1/6+o(1)}) = 1 + o(1)$ and similarly the other appearance of y are asymptotically negligible. Since $k \leq c \leq \frac{7}{3}k \leq \frac{5}{2}(k-1)$ the mentioned expression is at most

$$(1 + o(1)) \frac{en}{x} \left(\frac{5ex}{2n} \right)^{\frac{k-1}{2}} \leq (e/\theta_1 + o(1)) (5e\theta_1/2)^{\frac{k-1}{2}} < 0.8,$$

where the first inequality used the fact that $x \leq \theta_1 n$ and the second one holds for any $k \geq 15$. The occurrence of a bad set X in this range of x therefore has an overall probability of $O(0.8\sqrt{n})$.

Now take $\theta_2 = \frac{2}{7}$. To treat subsets X of size $\theta_1 n \leq x \leq \theta_2 n$ we will take advantage of the additional hypothesis about the edge set of H . For a given choice of V', X, Y such that all external neighbors of X are confined to Y (whose size we recall is $y = n^{1/3}$), there can be at most $|X| + |E_2| \leq |X| + n/\log \log n$ edges between X and $V' \setminus (X \cup Y)$ in G . The variable Γ counting the number of such edges has a law of

$\text{Bin}(x(n' - x - y), p)$, and in particular its mean is $x(n' - x - y)p \geq x(\frac{4}{5}n - \frac{2}{7}n - o(n))p > (c/2)x \geq 7x$ for large n , since $c \geq k \geq 15$. Thus, writing $\delta_n = 12/\log \log n$ so that $x + n/\log \log n \leq (1 + \delta_n)x$,

$$\begin{aligned} \mathbb{P}(\Gamma \leq (1 + \delta_n)x) &\leq (1 + \delta_n)x \binom{x(n' - x - y)}{(1 + \delta_n)x} p^{(1 + \delta_n)x} e^{-px(n' - x - y) + p(1 + \delta_n)x} \\ &\leq \left[(1 + o(1))ec(1 - x/n)e^{-c(n' - x)/n} \right]^x \leq \left[(1 + o(1))ce^{1 - c/2} \right]^x, \end{aligned}$$

where the last inequality plugged in the facts $n' \geq \frac{4}{5}n$ and $x \leq \frac{2}{7}n$. Armed with this expression we revisit (3.2) and revise it to incorporate the above probability, albeit at the cost of accounting the $\binom{n}{n'}$ possible choices for V' , so the probability of encountering the mentioned set X becomes at most

$$\begin{aligned} &\sum_{\theta_1 n \leq x \leq \theta_2 n} \sum_{\frac{4}{5}n \leq n' \leq n} \binom{n}{n'} \left[(1 + o(1)) \frac{en}{x} \left(\frac{ecx}{(k-1)n} \right)^{\frac{k-1}{2}} ce^{1 - c/2} \right]^x \\ &\leq \sum_{\theta_1 n \leq x \leq \theta_2 n} \left[(1 + o(1)) e^{2 + h(\frac{1}{5})\frac{n}{x}} \frac{n}{x} \left(\frac{ecx}{(k-1)n} \right)^{\frac{k-1}{2}} ce^{-c/2} \right]^x, \end{aligned}$$

where we used the fact that $\sum_{i \leq \alpha n} \binom{n}{i} \leq \exp(h(\alpha)n)$ with $h(x) = -x \log x - (1 - x) \log(1 - x)$ being the entropy function, and as argued before, the terms involving y in (3.2) are easily absorbed in the $(1 + o(1))$ -factor. Writing $\tilde{c} = c/(k-1)$ and $\tilde{x} = x/n$ (so that $\tilde{c} \leq \frac{5}{2}$ and $\theta_1 \leq \tilde{x} \leq \theta_2$) we can rearrange the base of this exponent in each summand and find that it is asymptotically equal to

$$e^2 \left(\tilde{x} e^{\frac{2}{k-3} h(\frac{1}{5})/\tilde{x}} \right)^{\frac{k-3}{2}} \left(\tilde{c} e^{1 - \tilde{c}} c^{\frac{2}{k-1}} \right)^{\frac{k-1}{2}} \leq e^2 \left(\tilde{x} e^{\frac{1}{6} h(\frac{1}{5})/\tilde{x}} \right)^{\frac{k-3}{2}} \left(\frac{5}{3} \right)^{\frac{k-1}{2}}$$

using the facts that $k \geq 15$, that $c^{\frac{2}{k-1}} \leq (\frac{5}{2}(k-1))^{\frac{2}{k-1}} \leq 5/3$ for $k \geq 15$ and that the function $t \mapsto te^{-t}$ with $t > 0$ has a global maximum at $t = 1$. Similarly, the function $t \mapsto te^{\alpha/t}$ with $\alpha, t > 0$ is increasing for $t \geq \alpha$ and as $\tilde{x} \geq \theta = \frac{2}{23} > \frac{1}{6} h(\frac{1}{5})$ this implies that the above expression is maximized at $\tilde{x} = \frac{2}{7}$, in which case it evaluates into

$$\frac{5}{3} e^2 \left(\frac{10}{21} e^{\frac{7}{12} h(\frac{1}{5})} \right)^{\frac{k-3}{2}} < 0.9$$

for any $k \geq 15$, and so the probability of encountering the mentioned set X in this range is $O(\exp(-an))$ for some absolute constant $a > 0$, which completes the proof. \blacksquare

In what follows, for a graph H and a subset A of its vertices let $c_o(H \setminus A)$ denote the number of components of odd size in the subgraph obtained by deleting the vertices of A from H . The above claims are already enough to imply a bound on $c_o(H \setminus A)$ for small subsets A , which will later be used as part of the proof of Lemma 2.3.

Corollary 3.3. *Fix $k \geq 15$ and let $G \sim \mathcal{G}(n, c/n)$ for some fixed c such that $\max\{k, 20\} \leq c \leq \frac{7}{3}k$. Then with probability $1 - o(n^{-3})$, for any induced subgraph $G' = (V', E') \subset G$ of minimum degree at least k on $|V'| \geq \frac{4}{5}n$ vertices and any $H = (V', E' \setminus (E_1 \cup E_2))$ such that E_1 is a matching, $|E_2| \leq n/\log \log n$ and $\delta(H) \geq k - 1$, we have $c_o(H \setminus X) \leq |X|$ for all $X \subset V'$ of size $|X| \leq \frac{3}{7}c^{-15/7}n$.*

Proof. Let $\theta = \frac{3}{7}c^{-15/7}$, and assume that the events stated in Claims 3.1 and 3.2 hold. In particular, the graph H under consideration satisfies that every set $S \subset V'$ of size at most θn has at least $3|S| - |S| = 2|S|$ external neighbors (deleting the matching E_1 can destroy at most $|S|$ external neighbors), and in addition it has at least $n^{1/3}$ external neighbors if its size lies between \sqrt{n} and $\frac{2}{7}n$.

Suppose first that $\log n \leq |X| \leq (\theta/2)n$ and let $\{C_1, \dots, C_s\}$ be all the connected components in $H \setminus X$ of size $|C_i| \leq \theta n$. Clearly $(N_H(C_i) \setminus C_i) \subset X$, but at the same time $|N_H(C_i) \setminus C_i| > 2|C_i|$ hence we must have $|C_i| < |X|/2$ for all i . Suppose that D is a minimal union of C_i 's such that $|D| > |X| - \frac{1}{3} \log n$. Since $|X| \geq \log n$ we have $|X| < \frac{3}{2}|D|$. Using that $|C_i| < |X|/2$, by the minimality of D , we also have $|D| \leq |X| - \frac{1}{3} \log n + |X|/2 < 3|X|/2 < \theta n$. Therefore $|N_H(D) \setminus D| \geq 2|D|$. However, at the same time $|N_H(D) \setminus D| \leq |X| < \frac{3}{2}|D|$ as there are no edges between distinct C_i, C_j , contradiction. Consequently, there cannot be more than $|X| - \frac{1}{3} \log n$ components of size at most θn (in fact, we showed that the union of all such components has cardinality at most $|X| - \frac{1}{3} \log n$). When added to at most $1/\theta$ components of $H \setminus X$ whose size exceeds θn we readily reach the required inequality $c_o(H \setminus X) \leq |X|$ in this case.

Next, consider the case $(\theta/2)n \leq |X| \leq \theta n$. Let us repeat the above argument, this time only collecting components of size at most $\log n$. Taking D to be a union of such components so that $\frac{2}{3}|X| - \log n < |D| \leq \frac{2}{3}|X|$ we see that $|N_H(D) \setminus D| \leq |X| \leq \frac{3}{2}|D| + \frac{3}{2} \log n$, which for n large enough contradicts the expansion assumption $|N_H(D) \setminus D| \geq 2|D|$. We deduce that there are at most $\frac{2}{3}|X|$ such components, and adding at most $n/\log n = o(|X|)$ components of size larger than $\log n$ gives $c_o(H \setminus X) \leq |X|$.

If $3 \leq |X| \leq \log n$ we observe that by our assumption there are no connected components of $H \setminus X$ of sizes between \sqrt{n} and $\frac{2}{7}n$. Consider the components C_1, \dots, C_s whose sizes are at most \sqrt{n} . Since there can be at most 3 components of size larger than $\frac{2}{7}n$, it follows that if $|\cup_i C_i| \leq |X| - 3$ then accounting for all such components results in $c_o(H \setminus X) \leq |X|$, as required. Otherwise, there exists a minimal union D of C_i 's such that $|D| \geq |X| - 2$. Being well in the range of the expansion hypothesis (this union has size at most $\sqrt{n} \log n$), the fact that $|X| \geq |N_H(D) \setminus D| \geq 2|D|$ implies that $|X| \leq 4$. However, in this case necessarily there are no small components in $H \setminus X$, since otherwise each C_i would have to be size at most $|X|/2$ by the expansion property, thus a vertex $v \in C_i$ would have degree at most $|C_i| - 1 + |X| \leq \frac{3}{2}|X| - 1 \leq 5$ (recall that $\delta(H) \geq k - 1 \geq 14$).

Finally, if $0 \leq |X| \leq 2$ then no component of $H \setminus X$ can have size less than $\frac{2}{7}n$ (recall that such components had to satisfy $|C_i| < |X|/2$ and are thus empty by the discussion in the previous paragraph). At the same time, the probability that two subsets A and B of size $\frac{2}{7}n$ each have no edges between them in G except for a matching (corresponding to E_1) is readily bounded by the probability that the degree into B of every vertex in A is at most 1, translating to an upper bound of

$$\binom{n}{\frac{4}{7}n} \binom{\frac{4}{7}n}{\frac{2}{7}n} \left(\left[\frac{2}{7}np + (1-p) \right] (1-p)^{\frac{2}{7}n-1} \right)^{\frac{2}{7}n} \leq \left(e^{\frac{7}{2}h(\frac{4}{7})} 2^2 \left(\frac{2}{7}c + 1 - o(1) \right) e^{-\frac{2}{7}c} \right)^{\frac{2}{7}n} \leq (0.99)^{2n/7},$$

where the last inequality holds for any $c \geq 20$. One can now repeat this calculation, this time taking into account the edges of E_2 . Since their number $(n/\log \log n)$ is negligible in comparison with the sizes of A and B , they contribute a factor of at most $\binom{(2n/7)^2}{n/\log \log n} p^{n/\log \log n} = e^{o(n)}$ in the above estimate and hence the conclusion remains valid. This concludes the proof. \blacksquare

A prerequisite for treating sets of larger size, which will also be useful to separately bound the size

of the 2-factor in G once we establish its existence, is the next estimate on the maximal number of vertex-disjoint cycles in G .

Claim 3.4. *Fix $c > 0$ and let $G \sim \mathcal{G}(n, c/n)$. Let Y count the maximal number of vertex-disjoint nontrivial cycles in G . Then $\mathbb{P}(Y \geq \frac{1}{2}n/\sqrt{\log n}) < e^{-\sqrt{n}}$ for any sufficiently large n .*

Proof. Let Y_0 be the maximal number of vertex-disjoint cycles in G such that each cycle is of length at most $L = 4\sqrt{\log n}$. Since $Y \leq Y_0 + n/L$ it suffices to show that $\mathbb{P}(Y_0 \geq n/L) \leq e^{-\sqrt{n}}$. Observe that

$$\mathbb{E}Y_0 \leq \sum_{s \leq L} \binom{n}{s} (s-1)! \left(\frac{c}{n}\right)^s \leq \sum_{s \leq L} c^s = e^{O(L)} = n^{o(1)} < n/(2L),$$

and consider the vertex-exposure martingale for Y_0 in $\mathcal{G}(n, c/n)$. Its Lipschitz constant is clearly equal to 1 since the addition of a new vertex to a graph retains every existing subset of disjoint cycles whereas it can increase the cycle count by at most 1. Hence, Hoeffding's inequality gives

$$\mathbb{P}(Y_0 \geq n/L) \leq \mathbb{P}(Y_0 - \mathbb{E}Y_0 \geq n/(2L)) \leq e^{-O(n/L^2)} < e^{-\sqrt{n}},$$

where the last inequality is valid for large enough n . ■

The final ingredient needed for proving Lemma 2.3 is the following claim, which verifies Tutte's condition for large sets and complements the range of sets that were covered by Corollary 3.3.

Claim 3.5. *Fix $k \geq 15$, let $G \sim \mathcal{G}(n, c/n)$ for $k \leq c \leq \frac{7}{3}k$ and let $G' = (V', E')$ be an induced subgraph of G with $\delta(G') \geq k$. Let $H = (V', E' \setminus (E_1 \cup E_2))$ where E_1 is a matching and $|E_2| \leq n/\log \log n$. Then with probability $1 - O(e^{-n/40})$ we have $c_o(H \setminus X) \leq |X|$ for all $X \subset V'$ such that $|X| \geq \frac{3}{7}c^{-15/7}n$.*

Proof. Set $x = |X|$ and consider first subsets of size

$$\frac{3n}{7c^{15/7}} \leq x \leq \frac{n}{c}. \quad (3.3)$$

Observe that by Claim 3.4 with probability at least $1 - O(e^{-\sqrt{n}})$ the maximal number of vertex disjoint cycles in G , and hence also in H , is at most $n/\sqrt{\log n}$. In particular, there can be at most that many connected components of $H \setminus X$ containing cycles. Disregard these components, as well as at most $n/\log \log n$ additional components incident to the edge set E_2 . Altogether we discarded $o(n)$ components, hence if $c_o(H \setminus X) > |X|$ for a linear X as above then we will be left with at least $\frac{15}{16}x$ odd components in $H \setminus X$ for large enough n . Let Y be a set comprised of an arbitrary leaf from each of these components (which are trees by construction). The vertices of Y have degree at least k in G' and as such they have at least $k - 2$ neighbors in X in the graph H (plus a single neighbor in their tree component of $H \setminus X$ and possibly one additional neighbor in E_1). The probability that there exists such a pair (X, Y) in G is at most

$$\sum_{x=\frac{3}{7}c^{-\frac{15}{7}}n}^{\frac{n}{c}} \binom{n}{x} \binom{\frac{15}{16}x}{x} \left[\mathbb{P}(\text{Bin}(x, c/n) \geq k-2) \right]^{\frac{15}{16}x} \leq \sum_{x=\frac{3}{7}c^{-\frac{15}{7}}n}^{\frac{n}{c}} \left[\frac{16}{15} \left(\frac{en}{x} \right)^{\frac{31}{15}} \mathbb{P}(\text{Bin}(x, c/n) \geq k-2) \right]^{\frac{15}{16}x}.$$

Plugging in the well-known estimate (see [3, 15]) that if Z is a binomial variable and $a > 0$ then $\mathbb{P}(Z - \mathbb{E}Z \geq a) \leq \exp(-\phi(a/\mathbb{E}Z)\mathbb{E}Z)$ for $\phi(x) = (1+x)\log(1+x) - x$, we get that

$$\begin{aligned} \frac{16}{15} \left(\frac{7}{3} e c^{15/7} \right)^{\frac{31}{15}} \mathbb{P}(\text{Bin}(x, c/n) \geq k-2) &\leq \frac{16}{15} \left(\frac{7}{3} e (7k/3)^{15/7} \right)^{\frac{31}{15}} (k-2)^{-(k-2)} e^{k-3} \\ &\leq 2e^6 k^5 \left(\frac{e}{k-2} \right)^{k-2} = 2e^{11} \left(1 + \frac{2}{k-2} \right)^5 \left(\frac{e}{k-2} \right)^{k-7} < 0.95, \end{aligned}$$

where the first inequality uses the fact that $c \leq \frac{7}{3}k$ and $xc/n \leq 1$, while the last inequality holds for $k \geq 15$. Combined with the previous inequality this implies that the probability of encountering some X violating Tutte's condition in H and with size as in (3.3) is at most $\exp(-c'n)$ for some fixed $c' > 0$.

We will now handle the range

$$\frac{n}{c} \leq x \leq (k-3)\frac{n}{c}. \quad (3.4)$$

Given a candidate subset X of size x for which $c_o(H \setminus X) > x$ we again consider the above defined set Y (comprised of an arbitrary leaf from each tree component of $H \setminus X$ not incident to any of the edges of E_2). Previously, we only used the fact that each $v \in Y$ has at least $k-2$ neighbors in X . Now we will use the fact that one can take Y to have size $(1-o(1))x$, and in addition Y is an independent set in H (its vertices fall in distinct components of $H \setminus X$). The probability that a given subset Y of size y in G spans precisely t edges is at most

$$\frac{\binom{y}{2}^t}{t!} p^t (1-p)^{\binom{y}{2}-t} \leq \frac{1}{t!} \left(\frac{\binom{y}{2}p}{1-p} \right)^t (1-p)^{\binom{y}{2}} =: \zeta_t,$$

which is maximized at $t = \binom{y}{2} \frac{p}{1-p} \geq (1-o(1))y/2$, with the last inequality due to $x \geq n/c$. Of course, the values of t we are considering cannot exceed $y/2$ as the edges of G in Y are only allowed to form a matching (recall that Y is not incident to E_2), thus by Stirling's formula

$$\sum_{t \leq y/2} \zeta_t \leq \left(\frac{y}{2} + 1 \right) \frac{1+o(1)}{\sqrt{\pi y}} \left(\frac{\binom{y}{2}p}{1-p} \frac{2e+o(1)}{y} (1-p)^{(y-1)} \right)^{(\frac{1}{2}-o(1))y} \leq ((e+o(1))p x e^{-px})^{(\frac{1}{2}-o(1))x}.$$

Combining this estimate with the number of choices for the sets X, Y and with the probability that each vertex of Y has at least $k-2$ neighbors in X , we arrive at the following upper bound on the event of encountering X, Y as above:

$$\sum_{x=n/c}^{(k-3)n/c} \left[(1+o(1)) \left(\frac{en}{x} \right)^2 \mathbb{P}(\text{Bin}(x, c/n) \geq k-2) \sqrt{epxe^{-px}} \right]^x, \quad (3.5)$$

where within each summand we could safely absorb all the $o(1)$ -terms in the exponents into the $o(1)$ -term of the leading constant since the base of each of these is uniformly bounded.

Notice that for $\ell \geq k-2$ we have

$$\frac{\mathbb{P}(\text{Bin}(x, p) = \ell + 1)}{\mathbb{P}(\text{Bin}(x, p) = \ell)} = \frac{x - \ell}{\ell + 1} \frac{p}{1-p} \leq (1+o(1)) \frac{px}{k-1},$$

which by our hypothesis on x is at most $(1+o(1))(k-3)/(k-1) < 1$ for large enough n . It follows that

$$\mathbb{P}(\text{Bin}(x, c/n) \geq k-2) \leq \frac{1+o(1)}{1 - \frac{px}{k-1}} \mathbb{P}(\text{Bin}(x, c/n) = k-2) \leq \frac{1+o(1)}{1 - \frac{px}{k-1}} \frac{(px)^{k-2}}{(k-2)!} e^{-px},$$

and plugging this in (3.5) gives an upper bound of

$$\sum_{x=n/c}^{(k-3)n/c} \left[(1+o(1)) \left(\frac{en}{x} \right)^2 \frac{1}{1 - \frac{px}{k-1}} \frac{(px)^{k-2}}{(k-2)!} e^{-px} \sqrt{epxe^{-px}} \right]^x.$$

Writing $z = px$ and $\tilde{c} = c/k$, we will denote the (asymptotic) base of the exponent above by

$$f_k(z) = \tilde{c}^2 e^{-\frac{3}{2}z + \frac{5}{2}} z^{k-\frac{7}{2}} \frac{k^2(k-1)}{(k-1-z)(k-2)!}.$$

With this notation we aim to bound $f_k(z)$ away from 1 for all $1 \leq z \leq k-3$ and $1 \leq \tilde{c} \leq \frac{7}{3}$. It is easy to verify that the local extrema of f_k are precisely the roots $z_k^* < z_k^{**}$ of the quadratic $3z^2 + (12-5k)z + (2k^2-9k+7)$, where z_k^* is a local maximum of f_k and z_k^{**} is a local minimum. Furthermore, since

$$z_k^{**} = \frac{5}{6}k - 2 + \frac{1}{6}\sqrt{k^2 - 12k + 60} > k-3$$

(the last inequality being valid for any $k > 0$) we can conclude that throughout the range $1 \leq z \leq k-3$ the function $f_k(z)$ is maximized at

$$z_k^* = \frac{5}{6}k - 2 - \frac{1}{6}\sqrt{k^2 - 12k + 60}.$$

Observe that $z_{k+1}^* \geq z_k^*$ and

$$\frac{2}{3}k - \frac{4}{3} \leq z_k^* \leq \frac{2}{3}k - 1.$$

In particular $z_{k+1}^* \leq \frac{2}{3}k - \frac{1}{3}$ and $0 \leq z_{k+1}^* - z_k^* \leq 1$.

Using the bound $z_k^* \leq \frac{2}{3}k - 1$ in the expression $1/(k-1-z)$ we immediately get that $f_k(z_k^*) \leq g_k(z_k^*)$ where

$$g_k(z) = 3\tilde{c}^2 e^{-\frac{3}{2}z + \frac{5}{2}} z^{k-\frac{7}{2}} \frac{k(k-1)}{(k-2)!}.$$

For ease of notation, let $z_0 = z_k^*$ and $z_1 = z_{k+1}^*$. We have

$$\frac{g_{k+1}(z_1)}{g_k(z_0)} = e^{-\frac{3}{2}(z_1-z_0)} \left(\frac{z_1}{z_0} \right)^{k-\frac{7}{2}} z_1 \frac{k+1}{(k-1)^2}.$$

Observe that

$$\left(\frac{z_1}{z_0} \right)^{k-\frac{7}{2}} \leq e^{\frac{z_1-z_0}{z_0}(k-\frac{7}{2})} \leq e^{\frac{3}{2} \frac{k-\frac{7}{2}}{k-2}(z_1-z_0)} \leq e^{\frac{3}{2}(z_1-z_0)},$$

and therefore

$$\frac{g_{k+1}(z_1)}{g_k(z_0)} \leq z_1 \frac{k+1}{(k-1)^2} \leq \frac{\frac{2}{3}(k-\frac{1}{2})}{k-1} \frac{k+1}{k-1} \leq \frac{2}{3} \left(1 + \frac{3}{k-1} \right) < \frac{5}{6},$$

where the last inequality is valid for any $k \geq 15$. It therefore suffices to show that $g_k(z_k^*)$ is bounded away from 1 for $k = 15$. Indeed, for $k = 15$ we have

$$z_{15}^* \in (10 - \frac{5}{4}, 10 - \frac{6}{5})$$

and substituting \tilde{c} by its maximal value of $\frac{7}{3}$ gives

$$g_{15}(z_{15}^*) \leq 49e^{-\frac{3}{2}(10-\frac{5}{4})+\frac{5}{2}} (10 - \frac{6}{5})^{15-\frac{7}{2}} \frac{70}{13!} < \frac{39}{40} < 1.$$

This establishes the desired statement for the range (3.4) provided that $1 \leq c/k \leq \frac{7}{3}$, that is, $c_o(H \setminus X) \leq |X|$ for all sets X whose sizes are between n/c and $(k-3)n/c$ except with probability exponentially small in n .

To conclude the proof it remains to handle $x \geq (k-3)n/c$. Observe that if $c \leq 2(k-3)$ then the range (3.4) goes up to $x \leq n/2$, beyond which $c_o(H \setminus X) \leq |X|$ trivially holds. We are thus left with the case $2(k-3) \leq c \leq \frac{7}{3}k$ and $(k-3)n/c \leq x \leq n/2$, where we claim that w.h.p. there will not exist a set Y as above which is an independent set in H . Indeed, the probability that $\mathcal{G}(n, c/n)$ will contain a set Y of size $(1 - o(1))(k-3)n/c \leq y \leq n/2$ where the edges form a matching (recall that Y is not incident with E_2) of size $\ell \leq y/2$ is at most

$$\begin{aligned} \binom{n}{y} \sum_{\ell \leq y/2} \binom{y}{2\ell} \frac{(2\ell)!}{\ell! 2^\ell} p^\ell (1-p)^{\binom{y}{2}-\ell/2} &\leq \sum_{\ell \leq y/2} \left[(e + o(1)) \frac{n}{y} \left(\frac{ey}{2\ell} \right)^{2\ell/y} \left(\frac{2\ell c}{en} \right)^{\ell/y} e^{-py/2} \right]^y \\ &\leq \sum_{\ell \leq y/2} \left[\frac{ec + o(1)}{k-3} \left(\frac{eyc}{4\ell} \right)^{\ell/y} e^{-(k-3)/2} \right]^y. \end{aligned}$$

It is easy to verify that the function $x \mapsto (ax)^{1/x}$ is decreasing in $(e/a, \infty)$ for any $a > 0$. In particular, for $a = ec/4 > e/2$ the maximum of $(ay/\ell)^{\ell/y}$ over all $y/\ell \geq 2$ is at most $\sqrt{2a}$, hence each of the summands in the right-hand-side above is at most

$$\left[\frac{ec + o(1)}{k-3} \sqrt{\frac{ec}{2}} e^{-(k-3)/2} \right]^y \leq \left[\frac{c^{3/2}e^3 + o(1)}{\sqrt{2}(k-3)} e^{-k/2} \right]^y \leq \left[5k^{3/2}e^{-k/2} \right]^y < \left(\frac{1}{5} \right)^{-y}.$$

where the second inequality uses that $c \leq \frac{7}{3}k$ and that $(\frac{7}{3})^{3/2} \frac{e^3}{\sqrt{2}(k-3)} < 5$ for $k \geq 15$. ■

We are now in a position to prove the main result of this subsection.

Proof of Lemma 2.3. The fact that G' is connected (as stated in Item (i) of the lemma) follows from Corollary 3.3 for a choice of $X = \emptyset$. Similarly, the expansion of sets of size at most n/c^3 (as stated in Item (ii) of the lemma) is an immediate consequence of Claim 3.1 (with plenty of room to spare, as $\frac{5}{4}c^{5/3} \leq c^2$ already for $c \geq 2$). It remains to establish the conclusion on the existence of a factor consisting of a path P and at most $n/\sqrt{\log n}$ cycles in G' .

The combination of Corollary 3.3 and Claim 3.5 shows that existence of a near-perfect matching (one that misses at most one vertex) in G' with probability at least $1 - o(n^{-3})$. Indeed, in the case where $|V'|$ is even, the existence of a perfect matching in G' follows from the celebrated Tutte condition [32] since $c_o(H \setminus A) \leq |A|$ for every $A \subset V'$. When $|V'|$ is odd, satisfying Tutte's condition for every $A \neq \emptyset$ will imply a near-perfect matching via Berge's formula (see, e.g., [22, §3.1.14]) since, as discussed above, G' is connected with probability $1 - o(n^{-3})$. Denoting this matching by M_1 we then re-apply Corollary 3.3 and Claim 3.5 on $H = (V', E' \setminus M_1)$ to extract another near-perfect matching M_2 with the same success probability. The factor formed by $M_1 \cup M_2$ is either a union of cycles (possibly with one isolated vertex which we count as a trivial cycle) or a union of cycles in addition to a single path. Moreover, the number of cycles is at most $n/\sqrt{\log n}$ except with exponentially small probability thanks to Claim 3.4, as required. ■

3.2 Properties of sparse random graphs beyond the core threshold

The main element needed for the proof of Lemma 2.4 is the expansion properties of sets in our random graph whose sizes exceed the one addressed in Claim 3.1.

Claim 3.6. *Let $G \sim \mathcal{G}(n, p)$ for $p = \lambda/n$ with $32 \leq \lambda \leq \sqrt{n}$. There exists an absolute constant $a > 0$ such that with probability at least $1 - O(e^{-an})$ we have $|N_G(X) \setminus X| > \frac{5}{2}|X|$ for every $X \subset V(G)$ such that $\frac{1}{2}\lambda^{-5/3}n \leq |X| \leq n/5$.*

Proof. Let $\theta_1 = \theta_1(n) = \frac{1}{2}\lambda^{-5/3}$ and $\theta_2 = 1/5$. Consider a potential subset X of x vertices such that $\theta_1 n \leq x \leq \theta_2 n$ and yet $|N_G(X) \setminus X| \leq b|x|$ for $b = \frac{5}{2}$. Letting $\tilde{x} = x/n$ for brevity, the probability that there exists such a set X is at most

$$\begin{aligned} \binom{n}{(b+1)x} \binom{(b+1)x}{x} (1-p)^{x(n-(b+1)x)} &\leq \left(\frac{en}{(b+1)x} \right)^{(b+1)x} e^{h(\frac{1}{b+1})(b+1)x} e^{-\frac{\lambda}{n}x(n-(b+1)x)} \\ &= \left(\frac{e^{1+h(\frac{1}{b+1})-\frac{\lambda}{b+1}} e^{\lambda\tilde{x}}}{b+1} \frac{1}{\tilde{x}} \right)^{(b+1)x} \end{aligned}$$

The unique extremum of $f(t) = e^{\lambda t}/t$ is a minimum at $t = 1/\lambda$, thus for $\theta_1 \leq \tilde{x} \leq \theta_2$ we have $f(\tilde{x}) \leq \max\{f(\theta_1), f(\theta_2)\}$. One can verify that

$$f(\theta_1) = 2\lambda^{5/3}e^{\lambda^{-2/3}/2} \leq 5e^{\lambda/5} = f(\theta_2)$$

for any $\lambda \geq 32$, and in particular the above probability is at most

$$\left(\frac{10}{7} e^{1+h(2/7)-(2/7)\lambda+\lambda/5} \right)^{\frac{7}{2}x} \leq \left(\frac{10}{7} e^{1+\log 2-\lambda/12} \right)^{\frac{7}{2}x} < 2^{-2x},$$

where in the last inequality is again valid for $\lambda \geq 32$. Summing over $O(n)$ values of $\theta_1 n \leq x \leq \theta_2 n$ now gives the desired result. \blacksquare

Observe that the claim above analyzed expansion in the entire random graph, while we will be mostly interested in the expansion properties of sets within the k -core. To this end we provide the following estimate on the size of core.

Claim 3.7. *Fix $k \geq 15$, let $G \sim \mathcal{G}(n, p)$ for $p = \lambda/n$ with $(7k-1)/3 \leq \lambda \leq \sqrt{n}$, and define*

$$\gamma = \gamma(k, \lambda) = (1 + \delta)e^k \left(\frac{\lambda}{k-1} \right)^{k-1} e^{-(1-\delta)\lambda} \quad \text{for } \delta = 10^{-3}. \quad (3.6)$$

Then $\gamma < 10^{-3}$ and $|V(G^{(k)})| > (1 - \gamma)n$ with probability at least $1 - O(e^{-an})$ for some absolute constant $a > 0$.

Proof. Consider a potential set S of $s = \lfloor \gamma n \rfloor$ vertices that do not belong to the k -core. We note that γ is decreasing in λ in the range $\lambda \geq k$ and so $\gamma(k, \lambda) \leq \gamma(k, \lambda_0)$ for $\lambda_0(k) = (7k-1)/3 = \frac{7}{3}(k-1) + 2$. At this value of λ we have

$$\gamma(k, \lambda_0) = (1 + \delta)e^{1-2(1-\delta)} \left(\frac{7e}{3} \left(1 + \frac{6}{7(k-1)} \right) e^{-(1-\delta)\frac{7}{3}} \right)^{k-1}, \quad (3.7)$$

and as the last factor is less than 1 for any $k \geq 4$, the entire expression is decreasing in k and so it is upper bounded by its value at $k = 15$, which, as one can easily verify, is strictly less than $10^{-3} = \delta$.

The iterative construction of the k -core via successively deleting vertices of degree less than k reveals that necessarily there are at most $(k-1)s$ edges in the cut between S and $V(G^{(k)})$. As the expected number of edges in this cut is $s(n-s)p \geq s\lambda/2 > s(k-1)$, the probability to encounter such a set S is at most

$$\begin{aligned} (s(k-1) + 1) \binom{n}{s} \binom{s(n-s)}{s(k-1)} p^{s(k-1)} (1-p)^{s(n-s)-s(k-1)} &\leq \left[(1+o(1)) \frac{en}{se^{p(n-s)}} \left(\frac{e(n-s)p}{k-1} \right)^{k-1} \right]^s \\ &= \left[\frac{e+o(1)}{\gamma e^{\lambda(1-\gamma)}} \left(\frac{e\lambda}{k-1} \right)^{k-1} \right]^s = \left[\frac{1+o(1)}{(1+\delta)e^{(\delta-\gamma)\lambda}} \right]^s < e^{-as} \end{aligned}$$

for some absolute constant $a > 0$, where the second equality substituted the definition of γ in (3.6) and the last inequality justified by the fact that $\gamma < \delta$. \blacksquare

Proof of Lemma 2.4. Let $H \sim \mathcal{G}(n, m)$ for $\frac{1}{6}(7k-1)n \leq m \leq n \log n$. Note that the edge density in H is $p \sim \lambda/n$ for $(7k-1)/3 \leq \lambda \leq 2\frac{\log n}{n}$, and in particular for $k \geq 15$ we see that λ satisfies the hypotheses of Claims 3.6 and 3.7. As per the conclusions of these claims, every set $X \subset V$ of size $\frac{1}{2}\lambda^{-5/3}n \leq |X| \leq n/5$ has $|N_H(X) \setminus X| > \frac{5}{2}|X|$, and $|V(H^{(k)})| \geq (1-\gamma)n$ for $\gamma < 0.001$ as defined in (3.6), except with probability $O(\exp(-an))$ for some absolute constant $a > 0$. Letting G be the result of adding $n/8$ uniformly chosen new edges to H , its edge density is asymptotically $(\lambda + \frac{1}{4})/n$, and appealing to Claim 3.1 we find that with probability $1 - o(n^{-3})$ every subgraph $F \subset G$ with minimum degree k has $|N_F(X) \setminus X| > 2|X|$ for any $X \subset V(F)$ of size at most $\frac{4}{5}\lambda^{-5/3}n$. Condition on these properties of H and G , from which we will now derive the required expansion properties of G' .

By definition, the graph G' has $V' = V(G^{(k)})$ as its vertex set, and its minimum degree is at least k , since every vertex $u \in V' \setminus V(H^{(k)})$ has the same degree in G' as it has in $G^{(k)}$ (at least k), whereas every $u \in V(H^{(k)})$ has degree at least k already within $H^{(k)} \subset G'$. Thus, thanks to our conditioning, every $X \subset V'$ such that $|X| \leq \frac{4}{5}\lambda^{-5/3}n$ satisfies $|N_{G'}(X) \setminus X| > 2|X|$.

It remains to treat sets $X \subset V'$ of size $\frac{4}{5}\lambda^{-5/3}n \leq |X| \leq n/5$. We claim that the proof will be concluded once we show that

$$\gamma(k, \lambda) \leq \frac{2}{5}\lambda^{-5/3} \quad \text{for any } \lambda \geq (7k-1)/3. \quad (3.8)$$

To see this, recall by our assumptions, for any set X in this size range, its set of external neighbors $Y = N_H(X) \setminus X$ satisfies $|Y| > \frac{5}{2}|X|$. Out of this set Y , all but at most γn vertices belong to V' (and hence belong to $N_{G'}(X) \setminus X$, as the induced subgraph of H on V' is a subgraph of G'). Thanks to (3.8) we have $\gamma n \leq |X|/2$, thus in particular $|N_{G'}(X) \setminus X| > 2|X|$, as desired.

Turning our attention to (3.8) and recalling the definition of γ from (3.6), we need to show that

$$(1+\delta) \left(\frac{e\lambda}{k-1} \right)^{k-1} e^{1-(1-\delta)\lambda} \cdot \frac{5}{2}\lambda^{5/3} \leq 1 \quad \text{for } \delta = 10^{-3}.$$

Since the function $x \mapsto x^{k-1+5/3} \exp(-(1-\delta)x)$ has a unique extremum in the form of a global maximum at $x = (k-1 + \frac{5}{3})/(1-\delta) < \frac{7}{3}k-1$, it suffices to establish the above inequality at

$\lambda_0 = \frac{7}{3}k - \frac{1}{3} = \frac{7}{3}(k-1) + 2$. Following (3.7), the left-hand-side of the sought inequality then becomes

$$(1+\delta) \cdot 5 \cdot 2^{2/3} \left(1 + \frac{7}{6}(k-1)\right)^{5/3} e^{1-2(1-\delta)} \left(\frac{7e}{3} \left(1 + \frac{6}{7(k-1)}\right) e^{-(1-\delta)\frac{7}{3}}\right)^{k-1}.$$

Denoting this expression by ξ_k , we need to show that $\xi_k \leq 1$ for all $k \geq 15$. We see that for any $k \geq 2$,

$$\frac{\xi_k}{\xi_{k-1}} = \left(1 + \frac{7}{7k-8}\right)^{5/3} \left(\frac{7e}{3} e^{-(1-\delta)\frac{7}{3}}\right) \left(1 + \frac{6}{7(k-1)}\right) \left(1 - \frac{6}{(k-1)(7k-8)}\right)^{k-2}.$$

Clearly each of the three factors that depend on k is decreasing in k (with the last of these being strictly less than 1 and raised to a power growing with k). In particular, for any $k \geq 5$ we have

$$\frac{\xi_k}{\xi_{k-1}} \leq \frac{\xi_5}{\xi_4} < 0.95 < 1.$$

Equivalently, ξ_k is strictly decreasing in k in the range $k \geq 5$, and so for any $k \geq 15$ we have

$$\xi_k < \xi_{15} < 0.95 < 1,$$

thus establishing (3.8), as required. ■

4 Packing edge-disjoint Hamilton cycles

In this section we prove Theorem 2. Since several of the ideas and techniques used here are rather similar to those applied beforehand, we will allow ourselves to be somewhat brief at times. Also, we will borrow extensively from the notation and terminology of the previous sections.

First of all, let

$$k_1 = \left\lfloor \frac{k-3}{2} \right\rfloor.$$

Similarly to Theorems 2.1 and 2.2, the following two theorems cover the critical regime (which is narrower here) and the supercritical regime, respectively.

Theorem 4.1. *Let k be a large enough constant, and let $G \sim \mathcal{G}(n, m)$ for $m = cn/2$, with c satisfying $k \leq c \leq k + 100\sqrt{k \log k}$. Then with probability $1 - o(n^{-2})$ either $G^{(k)}$ is empty or it contains a family of k_1 edge-disjoint Hamilton cycles.*

Theorem 4.2. *Let k be a large enough constant, and let $H \sim \mathcal{G}(n, m)$ for $m = cn/2$, with $c = c(n)$ satisfying $k + 99\sqrt{k \log k} \leq c \leq 2 \log n$. Let G be the graph obtained by adding n uniformly distributed new edges to H . Then the probability that $H^{(k)}$ contains a family of k_1 edge disjoint Hamilton cycles, and at the same time $G^{(k)}$ does not contain a family of k_1 edge-disjoint Hamilton cycles, is $o(n^{-2})$.*

The above two theorems, combined with the classical result of Bollobás and Frieze [8] that the random graph $G \sim \mathcal{G}(n, m)$ with $m = \frac{n}{2}(\log n + (k-1) \log \log n + \omega_n)$ and any diverging increasing sequence ω_n w.h.p. contains $\lfloor k/2 \rfloor$ edge-disjoint Hamilton cycles, clearly establishes Theorem 2.

In our proofs we prefer to work with the binomial random graph model $\mathcal{G}(n, p)$ instead of $\mathcal{G}(n, m)$; standard arguments about the essential equivalence of the models can be invoked to transfer our claims to the model $\mathcal{G}(n, m)$.

We start with the following lemma, guaranteeing typical local expansion in subgraphs of minimum degree at least 5 in sparse random graphs.

Lemma 4.3. *Let $G \sim \mathcal{G}(n, p)$ for $p = c/n$ with $c_0 \leq c(n) \leq 2 \log n$. If c_0 is large enough, then with probability $1 - o(n^{-2})$ every subgraph H of G of minimum degree $\delta(H) \geq 5$ has the following property: every subset $X \subset V(H)$ of size $|X| \leq n/c^{10}$ satisfies $|N_H(X) \setminus X| \geq 2|X|$.*

Proof. First observe that, for any given constant $C_0 > 0$, with probability $1 - o(n^{-2})$ every set $X \subset V(G)$ with $|X| \leq C_0$ vertices contains at most $|X| + 2$ edges. Indeed, the probability that there exists a subset violating this claim is at most

$$\sum_{i \leq C_0} \binom{n}{i} \binom{\binom{i}{2}}{i+3} p^{i+3} = O(n^{C_0} p^{C_0+3}) = o(n^{-2}).$$

Now we argue that typically in $\mathcal{G}(n, p)$ all larger sets are still fairly sparse. Specifically, we claim that there exists a large enough C_0 such that with probability $1 - o(n^{-2})$ every set $Z \subset V(G)$ of cardinality $C_0 \leq |Z| \leq 3n/c^{10}$ spans at most $1.2|Z|$ edges in G . Indeed, the probability that there exists a set Z of cardinality $|Z| = z$ violating this claim is at most

$$\begin{aligned} \sum_{z=C_0}^{3n/c^{10}} \binom{n}{z} \binom{\binom{z}{2}}{1.2z} p^{1.2z} &\leq \sum_{z=C_0}^{3n/c^{10}} \left[\frac{en}{z} \left(\frac{ezp}{2.4} \right)^{1.2} \right]^z \leq \sum_{z=C_0}^{3n/c^{10}} \left[5 \left(\frac{z}{n} \right)^{0.2} c^{1.2} \right]^z \\ &\leq \sum_{z=C_0}^{\log n} n^{-0.1z} + \sum_{z=\log n}^{3n/c^{10}} [10c^{-0.8}]^z = o(n^{-2}), \end{aligned}$$

for $c \geq c_0$ large enough and $z \geq C_0$ large enough.

Now take $C_0 \leq |X| \leq n/c^{10}$ and let $Y = N_H(X) \setminus X$. Since all degrees in H are at least 5, we deduce that $e_H(X, Y) \geq 5|X| - 2e_H(X)$, and thus $e_H(X \cup Y) = e_H(X) + e_H(X, Y) \geq 5|X| - e_H(X) \geq 3.8|X|$. If $|Y| < 2|X|$, then $|X \cup Y| < 3|X| \leq 3n/c^{10}$ and $e_H(X \cup Y) \geq 3.8|X| > 1.2|X \cup Y|$ — a contradiction.

For the sets X of constant size $|X| = x \leq C_0$, if $|N_H(X) \setminus X| < 2|X|$, then there is a set Y of cardinality $|Y| = 2x - 1$ with $N_H(X) \setminus X \subseteq Y$. Since all degrees in H are at least 5, and X spans at most $x + 2$ edges, it follows that $e_H(X \cup Y) \geq 5x - e(X) \geq 4x - 2$. The probability that there exists a pair X, Y with $|X| = x$ in $\mathcal{G}(n, p)$ is bounded from above by

$$\sum_{x \leq C_0} \binom{n}{x} \binom{n}{2x-1} \binom{\binom{3x-1}{2}}{4x-2} p^{4x-2} = O(n^{3x-1} p^{4x-2}) \leq n^{-x+1+o(1)} = o(n^{-2})$$

for $x \geq 4$. For $x = 3$, a simple union bound shows that the probability of the existence of a pair X, Y as above is at most $n^{-2x-1+\binom{x}{2}+o(1)} = o(n^{-2})$ and for $x = 1, 2$ the sought claim follows from the minimum degree assumption. \blacksquare

We will also need that typically in the supercritical regime the k -core contains a proportion of vertices rapidly approaching 1.

Lemma 4.4. *Let $G \sim \mathcal{G}(n, p)$ for $p = c/n$ with $k + 99\sqrt{k \log k} \leq c(n) \leq 2 \log n$. If k is large enough, then with probability $1 - o(n^{-2})$ the k -core satisfies $|G^{(k)}| \geq n(1 - c^{-200})$.*

Of course, there is nothing new in this lemma, and the estimate on the size of the k -core can easily be improved. We present its (straightforward) proof here mainly for the sake of completeness.

Proof. Let

$$t = n/c^{200}.$$

Construct the (possibly empty) k -core by peeling off repeatedly vertices of degree less than k . If this process lasts for t steps, then G contains a subset V_0 of $|V_0| = t$ vertices such that all degrees from V_0 to $V - V_0$ are less than k . For a given V_0 and a given $v \in V_0$, the degree of v to $V - V_0$ in $\mathcal{G}(n, p)$ is $\text{Bin}(n - t, p)$, and by Chernoff's inequality (see, e.g., [2]),

$$\begin{aligned} \rho &= \mathbb{P}(\text{Bin}(n - t, p) < k) \leq \exp\left(-\frac{((n - t)p - k)^2}{2(n - t)p}\right) \leq \exp\left(-\frac{((n - t)p - k)^2}{2np}\right) \\ &\leq \exp\left(-\frac{(c - c^{-199} - k)^2}{2c}\right). \end{aligned}$$

One can verify easily that for $c \geq k + 99\sqrt{k \log k}$ and large k it holds that

$$\frac{(c - c^{-199} - k)^2}{2c} \geq 300 \log c,$$

and thus $\rho \leq c^{-300}$. Hence the probability of the existence of a set V_0 as above is at most

$$\binom{n}{t} \rho^t \leq \left(\frac{en}{t}\right)^t \rho^t \leq (ec^{-100})^t = o(n^{-2})$$

for large enough c (or large enough k , as $c \geq k$). If no such set V_0 exists, then clearly the k -core of G has at least $n - t$ vertices, as required. \blacksquare

For the critical regime, the analog of Lemma 2.3 is now the following statement.

Lemma 4.5. *Fix k large enough and let $G \sim \mathcal{G}(n, m)$ with $m = cn/2$ and $k \leq c \leq k + 100\sqrt{k \log k}$. Then with probability $1 - o(n^{-2})$, for any $G' = (V', E') \subset G$ with minimum degree $\delta(G') \geq k$ such that $V' = V(G^{(k)})$ and $E(G^{(k)}) \setminus E'$ contains at most $n/\log \log n$ edges, the graph G' contains a $(2k_1)$ -factor.*

The proof of this lemma can be derived by repeating the arguments of Prałat, Verstraëte and Wormald [30] and of Chan and Molloy [10]. They proved that for large enough k , upon creation, the k -core $G^{(k)}$ with high probability either contains a $(k - 2)$ -factor or is $(k - 2)$ -factor-critical [30], or even contains a $(k - 1)$ -factor or is $(k - 1)$ -factor-critical [10]. (A graph G is t -factor-critical if for every $v \in V(G)$, the subgraph $G - v$ contains a t -factor.) Both these papers use Tutte's factor theorem to derive the likely existence of a required factor. As one can anticipate (and we indeed verified), altering the k -core slightly (by deleting at most $n/\log \log n$ edges) while preserving minimum degree k , do not influence the proof. We would also like to point out that the results in both papers [10, 30] used the assumption that $c \leq k + 2\sqrt{k \log k}$, but one can easily check that replacing 2 with any other constant will still work. Finally note that $2k_1 < k - 2$ and either $k - 2$ or $k - 1$ is even. Thus, Lemma 4.5 follows from the above results by the well-known fact that a $2s$ -factor contains a $2t$ -factor for all $t \leq s$.

Proof of Theorem 4.1 (sketch). We utilize the same approach, notation and terminology as in the proof of Theorem 2.1. We start by summarizing all the results that have been established in the proof of that theorem which we use here.

Let $\mathcal{S} = (e_1, \dots, e_m)$ be a uniformly chosen ordered subset of m edges out of the $\binom{n}{2}$ possible ones on the vertex set $V = [n]$, $m' = m - n/\log \log n$, and let G and H be the graphs on the vertex set V with edge sets $E(G) = \{e_1, \dots, e_m\}$, $E(H) = \{e_1, \dots, e_{m'}\}$. As $G \sim \mathcal{G}(n, m)$, if $G^{(k)}$ is empty then there is nothing left to prove. We therefore assume otherwise and aim to establish that $G^{(k)}$ contains a family of k_1 edge-disjoint Hamilton cycles with probability $1 - o(n^{-2})$. Given the above two graphs $H \subset G$ on the same vertex set V , define the partition of the vertex set of $G^{(k)}$, the k -core of G , into $\mathcal{B}_k = \mathcal{B}_k(G, H)$ and $\mathcal{C}_k = \mathcal{C}_k(G, H)$ so that \mathcal{B}_k consists of all vertices of $G^{(k)}$ having at least k neighbors in $V(G^{(k)})$ already in the subgraph H and $\mathcal{C}_k = V(G^{(k)}) \setminus \mathcal{B}_k$. Denote the indices of the subset of the final $m - m'$ edges having both endpoints in \mathcal{B}_k by \mathcal{J} and let $J = |\mathcal{J}|$. Recall that we have already established that upon conditioning on $\mathcal{B}_k, \mathcal{C}_k$ (such that $\mathcal{B}_k \cup \mathcal{C}_k \neq \emptyset$ by our assumption) as well as \mathcal{J} and all the edges $\mathcal{S}^* = \{e_j : j \notin \mathcal{J}\}$, the remaining J edges of \mathcal{S} are uniformly distributed over all edges missing from \mathcal{B}_k (that is, edges with both endpoints in \mathcal{B}_k that did not appear among $e_1, \dots, e_{m'}$). Moreover we also proved that $J \geq n/(100 \log \log n)$ with probability $1 - O(n^{-9})$.

Letting $F = (V, \mathcal{S}^*)$, as in the proof of Theorem 2.1 we have $V(F^{(k)}) = V(G^{(k)})$. Further note that $G' = F^{(k)}$ satisfies the conditions of Lemma 4.5 and therefore contains a $(2k_1)$ -factor. By a well-known fact from graph theory, this $(2k_1)$ -factor can be decomposed into k_1 edge-disjoint 2-factors $\Lambda_1, \dots, \Lambda_{k_1}$. Observe that, due to Claim 3.4, with probability $1 - o(n^{-2})$ each 2-factor Λ_i has at most $n/\sqrt{\log n}$ cycles. We have at our disposal a set \mathcal{J} of $J \geq n/(100 \log \log n)$ random edges to fall into the set \mathcal{B}_k . Split \mathcal{J} into k_1 sets $\mathcal{J}_1, \dots, \mathcal{J}_{k_1}$ of nearly equal sizes $|\mathcal{J}_i| \geq \lfloor J/k_1 \rfloor$ and use \mathcal{J}_i to convert the i -th 2-factor Λ_i into a Hamilton cycle, of course relying on the existing edges as the backbone. More specifically, we repeat k_1 iterations of the following process. Suppose we have already constructed Hamilton cycles H_1, \dots, H_{i-1} . Our goal is to convert Λ_i into a Hamilton cycle H_i using the edges of

$$G_i = G' - (H_1 \cup \dots \cup H_{i-1}) \cup (\Lambda_{i+1} \cup \dots \cup \Lambda_{k_1})$$

and the random edges from \mathcal{J}_i . Observe that the minimum degree of the relevant graph satisfies $\delta(G_i) \geq k - (2k_1 - 2) \geq 5$, thus making Lemma 4.3 applicable. By this lemma, all connected components of G_i are of size at least n/c^{10} and therefore G_i has at most c^{10} connected components. Moreover, since $|\mathcal{C}_k| \leq 2(m - m') = 2n/(\log \log n)$, at least $n/c^{10} - |\mathcal{C}_k| \geq n/c^{11}$ vertices of each component belong to \mathcal{B}_k . Thus, adding any random edge in \mathcal{B}_k will connect two of these components with probability at least c^{-22} . By standard concentration arguments, it is then easy to see that throwing in, say, $\log^2 n$ edges from \mathcal{J}_i will turn G_i into a connected graph with probability $1 - o(n^{-2})$.

Now we apply Lemma 2.6. Due to this lemma together with the expansion property guaranteed by Lemma 4.3, at any step G_i contains at least $n/(2c^{20})$ boosters. Moreover, the number of such boosters with both endpoints in \mathcal{B}_k is at least $n/(2c^{20}) - n|\mathcal{C}_k| \geq n/c^{21}$. Thus, the probability that a random edge falling into \mathcal{B}_k hits a booster is at least c^{-21} . Since Λ_i has only at most $n/\sqrt{\log n}$ cycles to begin with, hitting $n/\sqrt{\log n} + 1$ such boosters will transform it into a Hamilton cycle. Also note that we have not yet used the vast majority of the edges from \mathcal{J}_i , and can still add $|\mathcal{J}_i| - \log^2 n > n/(200 \log \log n)$ random edges to G_i . By adding all these edges, we collect a number of boosters which stochastically dominates a binomial random variable with parameters $n/(200 \log \log n)$ and c^{-21} . Therefore, by standard concentration arguments, we create the next Hamilton cycle H_i with probability at least $1 - o(n^{-2})$. Finally, we release the edges of Λ_i not used in H_i back into the graph, guaranteeing that in the next iteration too the minimum degree of the backbone graph is at least 5. ■

Proof of Theorem 4.2. Observe first that due to Lemma 4.4, the difference in sizes of the k -cores

$G^{(k)}$ and $H^{(k)}$ is relatively small:

$$|V(G^{(k)})| - |V(H^{(k)})| \leq n - |V(H^{(k)})| \leq n/c^{200}.$$

Let $V' = V(H^{(k)})$ and let R be the set of n random edges. For every edge $e = (u, v) \in R$ (uniformly distributed over missing edges in H) expose whether both u, v are in V' or not. If at least one of them is not in V' expose both endpoints of e . Otherwise put e into R' and note that R' is the set of random edges uniformly distributed over all missing edges in the induced graph of H on V' . Also note that since $|V'| \geq (1 - c^{-200})n$, the number of edges in R' is stochastically dominated by the binomial random variable with parameters n and $2/3$. Therefore with probability at least $1 - o(n^{-2})$ in the end of this process R' has at least $n/2$ edges. Moreover, this process reveals the set of vertices of the k -core of the graph G obtained from H by adding edges in R .

Fix a collection $\Lambda_1, \dots, \Lambda_{k_1}$ of edge-disjoint Hamilton cycles in $H^{(k)}$. We will use the set R' of random edges to convert them into edge-disjoint Hamilton cycles in $G^{(k)}$. To this end, split the set R' in k_1 parts R_1, \dots, R_{k_1} of sizes $|R_i| \geq n/(3k_1)$. We perform k_1 iterations of the following process. Suppose that for $i \in [k_1]$ we have already created Hamilton cycles H_1, \dots, H_{i-1} in $G^{(k)}$. Let

$$G_i = G^{(k)} - (H_1 \cup \dots \cup H_{i-1}) \cup (\Lambda_{i+1} \cup \dots \cup \Lambda_{k_1}).$$

By definition, the minimum degree of G_i is at least 5, enabling us to utilize Lemma 4.3. Due to this lemma, all connected components of G_i are of size at least n/c^{10} and therefore G_i has at most c^{10} connected components. Moreover, at least $n/c^{10} - n/c^{200} \geq n/c^{20}$ vertices of each component belong to V' . Thus adding any random edge in V' will connect two of these components with probability at least c^{-40} . Therefore, by standard concentration arguments, it is easily seen that throwing in (say) first $\log^2 n$ edges from R_i will turn G_i into a connected graph with probability $1 - o(n^{-2})$.

Next, as in the proof of Theorem 2.2, we add to Λ_i all vertices of $V(G^{(k)}) - V(H^{(k)})$ (at most n/c^{200} of them) to have a 2-factor with at most $n/c^{200} + 1$ cycles. We merge these cycles into one gradually, using the remaining $|R_i| - \log^2 n \geq n/(4k_1) > n/c^2$ of the edges of the R_i . By the expansion properties (Lemma 4.3) of G_i , we can derive from Lemma 2.6 that there are at least $n^2/(2c^{20})$ boosters in G_i . Moreover, all but n^2/c^{200} such boosters have both endpoints in V' . Hence, the probability that a new edge, uniform over all missing edges in V' , is a booster is at least c^{-21} . Hitting $n/c^{200} + 2$ such boosters will transform our 2-factor into a Hamilton cycle in G_i . Since the number of boosters that we encounter is stochastically dominated by a binomial variable $\text{Bin}(n/c^2, c^{-21})$, we deduce that with probability $1 - o(n^{-2})$ exposing random edges from R_i will create the next Hamilton cycle H_i in $G^{(k)}$. Finally, we release the edges of Λ_i not used in H_i back into the graph, guaranteeing that in the next iteration too the minimum degree of the backbone graph would be at least 5. \blacksquare

5 Open Problems

The main problem that remains open is, of course, whether the result of Theorem 1 (the Hamiltonicity of $G_t^{(k)}$ for all $t \geq \tau_k$ w.h.p. provided $k \geq 15$) can be pushed all the way down to $k \geq 3$. The bottleneck in our proof, as mentioned before, was finding a 2-factor in the k -core minus the unrevealed sprinkling edges. Due to the complexity of this graph, instead of appealing to Tutte's criterion for a 2-factor, we construct one using 2 disjoint perfect matchings. This appears to be possible at τ_k as long as $k \geq 7$.

By taking $k \geq 15$ we were able both to simplify the arguments used to recover this 2-factor, as well as to guarantee that $G_t^{(k)}$ would remain Hamiltonian for all $t \geq \tau_k$ w.h.p. It is thus quite plausible that, when paired with a more refined analysis, the present framework should produce an improved bound on k . However, a new idea seems to be needed to take it down to $k = 3$.

For the packing problem, while we are only one Hamilton cycle short of the conjectured packing of $\lfloor (k-1)/2 \rfloor$ edge-disjoint Hamilton cycles in the k -core, the present methods appear to be incapable of bridging this small gap. In particular, in our cycle-by-cycle extraction argument it is crucial to have minimum degree 5 at every stage, so as to guarantee sufficient expansion of small sets (as given by Lemma 4.3 and used by Lemma 2.6). Once we extract $\lfloor (k-3)/2 \rfloor$ Hamilton cycles from the k -core, we will be left with a subgraph of minimum degree 3, and this by itself is not enough to ensure the required expansion (e.g., consider a set of degree-3 vertices forming a cycle). In conclusion, while new ideas are needed to fully settle this conjecture, our result provides a fairly firm evidence for its validity.

Acknowledgment. A major part of this work was carried out when the first and the third authors were visiting Microsoft Research at Redmond, WA. They would like to thank the Theory Group at Microsoft Research for hospitality and for creating a stimulating research environment.

References

- [1] M. Ajtai, J. Komlós, and E. Szemerédi, *First occurrence of Hamilton cycles in random graphs*, Cycles in graphs (Burnaby, B.C., 1982), North-Holland Math. Stud., vol. 115, North-Holland, Amsterdam, 1985, pp. 173–178.
- [2] N. Alon and J. H. Spencer, *The probabilistic method*, 3rd ed., John Wiley & Sons Inc., 2008.
- [3] G. Bennett, *Probability inequalities for the sum of independent random Variables*, Journal of the American Statistical Association **57** (1962), no. 297, 33–45.
- [4] B. Bollobás, *The evolution of sparse graphs*, Graph theory and combinatorics (Cambridge, 1983), Academic Press, London, 1984, pp. 35–57.
- [5] B. Bollobás, *Random graphs*, 2nd ed., Cambridge University Press, Cambridge, 2001.
- [6] B. Bollobás, C. Cooper, T. I. Fenner, and A. M. Frieze, *Edge disjoint Hamilton cycles in sparse random graphs of minimum degree at least k* , J. Graph Theory **34** (2000), no. 1, 42–59.
- [7] B. Bollobás, T. I. Fenner, and A. M. Frieze, *Hamilton cycles in random graphs of minimal degree at least k* , A tribute to Paul Erdős, Cambridge Univ. Press, Cambridge, 1990, pp. 59–95.
- [8] B. Bollobás and A. M. Frieze, *On matchings and Hamiltonian cycles in random graphs*, Random graphs '83 (Poznań, 1983), North-Holland Math. Stud., vol. 118, North-Holland, Amsterdam, 1985, pp. 23–46.
- [9] J. Cain and N. Wormald, *Encores on cores*, Electron. J. Combin. **13** (2006), no. 1, Research Paper 81, 13 pp. (electronic).
- [10] S. O. Chan and M. Molloy, *$(k+1)$ -Cores Have k -Factors*, Combinatorics, Probability and Computing **21** (2012), no. 6, 882–896.
- [11] V. Chvátal, *Almost all graphs with $1.44n$ edges are 3-colorable*, Random Structures Algorithms **2** (1991), no. 1, 11–28.

- [12] C. Cooper, *The cores of random hypergraphs with a given degree sequence*, Random Structures Algorithms **25** (2004), no. 4, 353–375.
- [13] P. Erdős and A. Rényi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **5** (1960), 17–61.
- [14] A. Frieze, *On a greedy 2-matching algorithm and Hamilton cycles in random graphs with minimum degree at least three*, Random Structures Algorithms, to appear.
- [15] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30.
- [16] S. Janson and M. J. Luczak, *A simple solution to the k -core problem*, Random Structures Algorithms **30** (2007), no. 1–2, 50–62.
- [17] S. Janson and M. J. Luczak, *Asymptotic normality of the k -core in random graphs*, Ann. Appl. Probab. **18** (2008), no. 3, 1085–1137.
- [18] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.
- [19] J. Komlós and E. Szemerédi, *Limit distribution for the existence of Hamiltonian cycles in a random graph*, Discrete Math. **43** (1983), no. 1, 55–63.
- [20] A. Korshunov, *Solution of a problem of Erdős and Rényi on Hamilton cycles in non-oriented graphs*, Soviet Math. Dokl. **17**, 760–764.
- [21] M. Krivelevich, E. Lubetzky, and B. Sudakov, *Hamiltonicity thresholds in Achlioptas processes*, Random Structures Algorithms **37** (2010), no. 1, 1–24.
- [22] L. Lovász and M. D. Plummer, *Matching theory*, AMS Chelsea Publishing, Providence, RI, 2009.
- [23] T. Łuczak, *On matchings and Hamiltonian cycles in subgraphs of random graphs*, Random graphs '85 (Poznań, 1985), North-Holland Math. Stud., vol. 144, North-Holland, Amsterdam, 1987, pp. 171–185.
- [24] T. Łuczak, *Size and connectivity of the k -core of a random graph*, Discrete Math. **91** (1991), no. 1, 61–68.
- [25] T. Łuczak, *Sparse random graphs with a given degree sequence*, Random graphs, Vol. 2 (Poznań, 1989), Wiley-Intersci. Publ., Wiley, New York, 1992, pp. 165–182.
- [26] M. Molloy, *A gap between the appearances of a k -core and a $(k + 1)$ -chromatic graph*, Random Structures Algorithms **8** (1996), no. 2, 159–160.
- [27] M. Molloy, *Cores in random hypergraphs and Boolean formulas*, Random Structures Algorithms **27** (2005), no. 1, 124–135.
- [28] B. Pittel, J. Spencer, and N. Wormald, *Sudden emergence of a giant k -core in a random graph*, J. Combin. Theory Ser. B **67** (1996), no. 1, 111–151.
- [29] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Math. **14** (1976), no. 4, 359–364.
- [30] P. Prałat, J. Verstraëte, and N. Wormald, *On the threshold for k -regular subgraphs of random graphs*, Combinatorica **31** (2011), no. 5, 565–581.
- [31] O. Riordan, *The k -core and branching processes*, Combin. Probab. Comput. **17** (2008), no. 1, 111–136.
- [32] W. T. Tutte, *The factorization of linear graphs*, J. London Math. Soc. **22** (1947), 107–111.